

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL

FACULDADE DE BIOCÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

**EVOLUÇÃO E MODELAGEM MOLECULAR DO
FATOR NEUROTROFICO DERIVADO DO CÉREBRO (BDNF) EM
MAMÍFEROS**

FABÍOLA SALENE MATTEI

Porto Alegre
2010

**EVOLUÇÃO E MODELAGEM MOLECULAR DO
FATOR NEUROTROFICO DERIVADO DO CÉREBRO (BDNF) EM
MAMÍFEROS**

FABÍOLA SALENE MATTEI

Dissertação apresentada ao Programa
de Pós-Graduação em Biologia Celular e
Molecular da Faculdade de Biociências
da Pontifícia Universidade Católica do
Rio Grande do Sul.

Orientador: Eduardo Eizirik

Co-orientador: Osmar Norberto de Souza

Porto Alegre
2010

"I believe that the study of science, the learning of the scientific method by all people, will ultimately help the people of the world in the solution of our great social and political problems."

Linus Pauling

AGRADECIMENTOS

Bernadete Baseggio Mattei

João Maximino Mattei

Rodolfo Devino Baseggio Mattei

“Sou parte da história de vocês, na imensa narrativa de suas vidas. Portanto esta sub-parte Pós-Graduação, também é, concomitantemente, esforço e dedicação de vocês! Amo vocês.”

Eduardo Eizirik

Osmar Norberto de Souza

“Ao conhecimento brilhante e orientações nestes dois anos. Levo na minha bagagem estes conhecimentos científicos distintos, ambos abrangentes e, sobretudo o crescimento pessoal.”

LISTA DE FIGURAS E TABELA

- Figura 1.** Alinhamento das seqüências de aminoácidos das Neurotrofinas.....8
- Figura 2.** Anotação da seqüência protéica do BDNF em humanos.....12
- Figura 3.** Esquema da estrutura neurotrófica madura típica.....14
- Tabela 1.** Modelos protéicos cristalográficos disponíveis.....15
- Figura 4.** Interação do BDNF com outras proteínas.....17

LISTA DE SIGLAS

- antiBDNF:** BDNF opposite strand (fita oposta do BDNF)
- BDNF:** Brain derived neurotrophic factor (Fator neurotrófico derivado do cérebro)
- BLAST:** Basic local alignment search tool (Ferramenta básica de alinhamento local)
- IDRs:** Intrinsic disorder regions (Regiões intrínsecas desordenadas)
- LTD:** Long-term depression (Depressão de longa duração)
- LTP:** Long-term potentialization (Potencialização de longa duração)
- mtDNA:** Mitochondrial DNA (DNA mitocondrial)
- MeCP2:** Methyl CpG binding protein 2 (Proteína de ligação Metil CpG 2)
- NATs:** Natural antisense transcripts (Transcritos antisensos naturais)
- NGF:** Nerve growth factor (Fator de crescimento do nervo)
- NT-3:** Neurotrophin 3 (Neurotrofina 3)
- NT-4:** Neurotrophin 4 (Neurotrofina 4)
- p75^{NTR}:** p75 neurotrophic receptor (receptor neurotrófico p75)
- matBDNF:** C-terminal form of BDNF (Forma C-terminal do BDNF)
- proBDNF:** N-terminal domain of BDNF (Domínio N-terminal do BDNF)
- proNGF:** N-terminal domain of NGF (Domínio N-terminal do NGF)
- rRNA:** Ribosomal RNA (RNA ribossomal)
- Trk:** Tyrosine kinase receptor (receptor tirosina cinase)
- TrkB:** Tyrosine kinase receptor B (receptor tirosina cinase B)

RESUMO

O fator neurotrófico conhecido como BDNF tem atraído considerável atenção na literatura científica, pelo seu envolvimento em processos cruciais de desenvolvimento e regulação do sistema nervoso. Entretanto, muitos aspectos desta molécula são ainda desconhecidos, enquanto outros foram investigados apenas em humanos e organismos-modelo. Neste contexto, ferramentas de bioinformática podem ser muito úteis para realizar análises comparativas enfocando aspectos evolutivos, estruturais e funcionais. Este projeto objetivou uma caracterização de padrões de conservação e variação desta molécula através da comparação das seqüências codificantes deste gene em diferentes linhagens de mamíferos, assim como uma investigação de aspectos estruturais através da construção de um modelo 3D do BDNF consistente com dados experimentais. Os resultados revelam padrões bastante interessantes de conservação e variação de seqüência em diferentes linhagens, incluindo fortes evidências da ocorrência de seleção natural negativa (indicando restrição à mudança devido à relevância funcional) tanto no domínio maduro como no domínio pro. Além disso, observou-se evidência de seleção positiva (indução de mudanças possivelmente adaptativas) em diferentes linhagens, em todos os casos afetando diretamente o domínio pro. Estes resultados indicam que o domínio pro apresenta grande relevância funcional, e salientam a importância de focar esforços de investigação experimental nesta porção do BDNF.

Palavras-Chave: ProBDNF, forma homodimérica, seleção natural, mamíferos.

ABSTRACT

The neurotrophic factor known as BDNF has attracted considerable attention in the scientific literature, due to its involvement in critical processes related to the development and regulation of the nervous system. However, many aspects of this molecule are still unknown, while others have only been investigated in humans and model organisms. In this context, bioinformatic tools can be very useful to perform comparative analyses focusing on evolutionary, structural and functional aspects of this protein. This study aimed to characterize the patterns of conservation and variability in this molecule through the comparison of coding sequences in different mammalian lineages, as well as to investigate structural aspects by constructing a novel 3D model of a BDNF homodimer that is consistent with experimentally validated data. Our results revealed very interesting patterns of conservation and variability in different lineages, including strong evidence for the occurrence of negative natural selection (indicating constraints to evolutionary change implying functional relevance) in both the mature and the pro domains. In addition, we observed evidence of positive selection (induction of changes that are likely adaptive) in different lineages, in every case directly affecting the pro domain. These results indicate that the pro domain presents great functional relevance, and highlight the importance of focusing research efforts on this portion of BDNF.

Key-words: ProBDNF, homodimeric form, natural selection, mammals.

ÍNDICE

Agradecimentos.....	i
Lista de Figuras e Tabelas.....	ii
Lista de Siglas.....	iii
Resumo.....	iv
Capítulo 1 – Introdução e Objetivos.....	6
1. 1. BDNF como membro da família das neurotrofinas.....	6
1.2. Estrutura e expressão do gene <i>BDNF</i>	8
1.3. Proteína BDNF.....	9
1.4. Sinalização mediada pela proteína BDNF.....	15
1.5. Alinhamento de seqüências e caracterização molecular evolutiva.....	18
1.6. Modelagem molecular por homologia.....	21
1.7. Objetivos.....	23
Capítulo 2 – Artigo Científico.....	24
2.1. Informação suplementar.....	32
Capítulo 3 – Considerações Finais.....	58
Referências.....	60

Capítulo 1 – Introdução e Objetivos.

1. 1. BDNF como membro da família das neurotrofinas.

O Fator neurotrófico derivado do cérebro (BDNF) é um membro da família de proteínas homólogas conhecidas como neurotrofinas, e tem um papel central no desenvolvimento, fisiologia, e patologia do sistema nervoso; como também em processos relacionados à plasticidade cerebral como a memória, e o aprendizado (*Yamada et al., 2002*). Neurotrofinas, que em mamíferos, incluem o Fator de crescimento nervoso (NGF), a Neurotrofina 3 (NT-3), a Neurotrofina 4 (NT-4), e o BDNF, têm sido caracterizadas em rotas funcionais distintas, porém paralelas (*Hallböök, 1999*). O BDNF foi o segundo fator trófico a ser identificado e caracterizado (*Barde et al., 1982*), antecedido pelo NGF; esse último apresenta-se como molécula-modelo para os estudos neurotróficos. Os primeiros estudos que abrangeram a filogenia molecular envolvendo as neurotrofinas relataram uma seqüência de eventos para a formação dos genes neurotróficos: duas duplicações lideraram a formação desta família; após um episódio inicial de duplicação, um dos genes descendentes deu origem ao par *BDNF* e *NT-4*, enquanto o outro ao par *NT-3* e *NGF* (*Hallböök, 1999; Lanave et al., 2007*). Só após o segundo evento de duplicação houve a diferenciação dos pares, caracterizada pela especialidade individual de cada gene neurotrófico. Do mesmo modo, o alinhamento múltiplo das seqüências de aminoácidos, revela cerca de 40% de similaridade entre os membros das neurotrofinas (*Figura 1*), e estudos iniciais demonstraram que as regiões conservadas e variáveis nestas seqüências estão agrupadas em blocos (*Robinson et al., 1995; Butte et al., 1998; Robinson et al., 1999*).

```

gi|128162|an|P01139.2|NGF_MOUSE      MSMLFYTLITAFILIGVQAEFYTDSDNVPEGDSVPEAHWIKLQHSALDTRR 50
gi|114900|an|P23560.1|BDNF_HUM      MTILFLTMVISYFGCMKAAPMKEANIRGQGLAYPGVRIHGTLESVNGPK 50
gi|128581|an|P20783.1|NTF3_HUM      MSILFYVIFLAYLRGIQGNMQRSLPEDSLNSLIKLIKQADILKKNLSK 50
gi|462741|an|P34130.1|NTF4_HUM      -----MLPLPSCSLPILLFLPLSPVPIESQPPPTSLPP 33
                                     . .:

gi|128162|an|P01139.2|NGF_MOUSE      ARSAPTAPIAARVTGQTRN-----ITVDPRLFKRRRLH 83
gi|114900|an|P23560.1|BDNF_HUM      AGSRGLTSLADTFEHVIEELLEDDQK-----VRPNEENKADLY 90
gi|128581|an|P20783.1|NTF3_HUM      QMVDVKENYQSTLPKAEAPREPERGGPAKSAFQPVIAMDTELLRQQRRYN 100
gi|462741|an|P34130.1|NTF4_HUM      FLAP-----EWDLL 42
                                     .

gi|128162|an|P01139.2|NGF_MOUSE      SPRVLFSTQPPPTSSDILDLDLQAHGIIPFNTRHRSKR--SSTHFVFMHG 131
gi|114900|an|P23560.1|BDNF_HUM      TSRVMLSSQVPLEPPLFLLEEYKNYLDAAANMSMRVRR----HSDPARRG 136
gi|128581|an|P20783.1|NTF3_HUM      SPRVLLSDSTPLEPPPLYLMEDYVGSFVAVNRTSRKR---YAEHKSHRG 147
gi|462741|an|P34130.1|NTF4_HUM      SPRVLSRGAPAGPPLLELLEAGAFRESAGAFAPNRSRRVSETAPASRRG 92
:.*:.* * .. : : * : * : *
                                     .

gi|128162|an|P01139.2|NGF_MOUSE      EFSVCDSESVVVG--DKTITATDIKQKEVTVLAEVNINNSV-FRQYFFETK 178
gi|114900|an|P23560.1|BDNF_HUM      ELSVCDSESEWVIAADKKTAVDMSSGGIVTVLEKVPVSKGQ-LKQYFYETK 185
gi|128581|an|P20783.1|NTF3_HUM      EYSVCDSESLWVI--DKSSAIDIRGHQVTVLGEIKTGNF-VKQYFYETR 194
gi|462741|an|P34130.1|NTF4_HUM      ELAVCDVAVSGWVI--DRRTAVDLRGREVEVLGEVPAAGGSPLRQYFFETR 140
* :***: * ** * : * * : * * * * : . .:***:***:
                                     .

gi|128162|an|P01139.2|NGF_MOUSE      CRASNFVESG-----CRGIDSKHWNSYCTITHTFVKALITD-EKQAAN 220
gi|114900|an|P23560.1|BDNF_HUM      CNPMGYTKEG-----CRGIDKRHWNSQCRTIQSYVRLIMDSKRRIGW 228
gi|128581|an|P20783.1|NTF3_HUM      CKEARFVKNG-----CRGIDDKHWNSQCRTISQTVVRLITSENNKLVGW 237
gi|462741|an|P34130.1|NTF4_HUM      CKADNAEEGGFGAGGGGCRVDRRHVWVSECKAKQSYVRLITADAQGRVVGW 190
* . : * * * * * : * * * * : . : * * * * : *
                                     .

gi|128162|an|P01139.2|NGF_MOUSE      RFIRIDTACVCLSRKATRRG 241
gi|114900|an|P23560.1|BDNF_HUM      RFIRIDTSCVCTLTIKRGR-- 247
gi|128581|an|P20783.1|NTF3_HUM      RWIRIDTSCVCLSRKIGRI- 257
gi|462741|an|P34130.1|NTF4_HUM      RWIRIDTACVCTLLSRIGRA- 210
*:*****:***.* : *

```

Figura 1. A Figura representa as quatro seqüências protéicas da família das neurotrofinas utilizadas para os experimentos de raios-X. **Legenda:** Precursor NGF-beta (241 aa); Precursor BDNF (247 aa); Precursor NT-3 (257 aa); Precursor NT-4 (210 aa), com suas respectivas identificações nos bancos de dados (gi: *sequence identifiers*; an: *accession number*); * região nas seqüências alinhadas de identidade de aminoácido; ; região nas seqüências alinhadas de similaridade de aminoácidos; _ região nas seqüências alinhadas de aminoácidos não-relacionados. O formato das seqüências utilizado para o alinhamento foi *Pearson*, e a matrix *blosum*.

As ações celulares das neurotrofinas são mediadas através de dois tipos de receptores: um receptor tirosina cinase (*Trk*) de alta afinidade, e um receptor pan-neurotrófico de baixa afinidade ($p75^{NTR}$). A comparação das análises filogenéticas entre as neurotrofinas e os seus receptores *Trks* identificou a co-evolução entre estes *locos*, fato que aumenta a especificidade entre as interações receptor-ligante (*Halböök et al., 1998*). Embora os efeitos tróficos de longa duração (potenciação e depressão) das neurotrofinas dependam da regulação gênica, os receptores ativados pelos seus membros também exercem uma ampla gama de ações rápidas, incluindo efeitos morfogenéticos e quimiotrópicos nos neurônios em desenvolvimento, assim como a modulação da excitabilidade neural, e a transmissão sináptica (*Poo, 2001*).

1.2. Estrutura e expressão do gene *BDNF*.

Em relação ao contexto cromossômico, o *BDNF* (*Gene ID: 627*) localiza-se no braço curto do cromossomo 11 humano perto de dois pseudogenes, ambos inferidos: ribosomal protein S25 pseudogene 1 (*Balasubramanian et al., 2009*) e chromobox homolog 3 pseudogene 1 (*Park et al., 1998*); este último envolvido na estrutura e organização da cromatina. O primeiro gene, localizado a 49kb *downstream* ao *BDNF*, é o *LIN-7* (*GeneID: 55327*), que codifica para uma proteína pequena descoberta em *Caenorhabditis elegans*. Lanktree e colaboradores (2008) encontraram associação entre polimorfismos analisando o *BDNF* e o *LIN-7* com a desordem de déficit de atenção / hiperatividade (ADHD). Já o gene localizado *upstream* ao *BDNF* é um membro citosólico da classe de proteínas heat shock, conhecido como *HSP90AA2* (*GeneID: 3324*); essa classe se associa com outras chaperonas, e tem um papel importante no enovelamento de novas proteínas, e/ou na estabilidade de proteínas desnaturadas depois de condições de estresse (*Ammirante et al., 2007*)

Análises comparativas iniciais incentivaram o estudo com este gene, e ainda apresentam importância para a caracterização sequencial deste gene em inúmeras outras espécies, destacando os mamíferos, foco desse estudo, com mais de 203 sequências descritas, e depositadas em bancos de dados. Atualmente o estudo da estrutura gênica e padrões de expressão em roedores e primatas estão sendo explorados. Neste contexto, a estrutura do gene *BDNF*, estudada em roedores (*Aid, et al., 2007*), demonstrou que camundongo e rato têm oito exons 5', e um exon 3', sendo que este último codifica para a proteína *BDNF* madura. *Pruunsild e colaboradores*, no mesmo ano, realizaram um amplo estudo do gene *BDNF* em humanos, e constataram a presença de onze exons, e nove promotores funcionais que determinaram à expressão tecido-específica dos transcritos, como predominantemente localizada no Sistema Nervoso Central; também, neste estudo,

foram detectados transcritos no Sistema Nervoso Periférico - assim como associado a um amplo padrão de expressão, por exemplo, no coração, rim, músculo, próstata, placenta, estômago e, testículo - similar ao padrão de expressão do *mRNA* observado para roedores.

1.3. Proteína BDNF.

A proteína BDNF foi originalmente purificada a partir de cérebro de porco, representando o segundo fator neurotrófico a ser purificado, depois do NGF, datando quase 30 anos (*Barde, 1982*). O estudo comparou esta proteína purificada com o NGF – fator trófico até então bem caracterizado – e constatou propriedades antigênicas e funcionais diferentes. O nível de expressão da proteína BDNF em humanos é regulado durante o desenvolvimento, mas persiste em muitas partes do cérebro adulto, sendo mais abundante no sistema nervoso, em comparação a outros tecidos, no referido período (*Katoh-Semba et al., 1997*). A clonagem e os estudos de expressão inicial desse gene (*Leibrock et al., 1989*) contribuíram para o interesse de inúmeros trabalhos posteriores, que se dedicaram à análise da expressão protéica em diversos tecidos, e a traçar prováveis rotas para sua sinalização, visto sua grande importância para a pesquisa em neurociências.

Entre as distintas isoformas identificadas em humanos, àquela denominada como isoforma *a* perfaz a maioria das pesquisas. Esta isoforma constitui-se em uma pré-proteína (*Uniprot p23560*), de 247 aminoácidos (*Figura 2A*), que é transportada através de vesículas para o retículo endoplasmático. A forma longa do proBDNF (sem o peptídeo sinal de direcionamento) (*Figura 2B*) possui dois motivos de clivagem; o primeiro origina uma forma curta do proBDNF (*Figura 2C*), e o segundo uma forma madura (*Figura 2D*), ambas independentes (*Mowla et al., 2001; Lanave et al., 2007*). Em relação à

primeira clivagem, a forma longa do pro-BDNF é processada intracelularmente através da proteinase SKI-1, na qual reconhece o motivo consenso $^{54}\text{RGLT}\downarrow\text{SL}^{59}$ de resíduos não-básicos, e cliva exatamente no aminoácido ^{57}T , originando a forma curta do proBDNF. (Seidah *et al.*, 1999). A segunda clivagem, não-dependente do primeiro processamento, ocorre extracelularmente através da enzima plasmina em uma região de resíduos básicos, no motivo consenso $^{125}\text{RVRR}^{128}\downarrow$ (Lee *et al.*, 2001; Pang *et al.*, 2004; Gray & Ellis, 2008), para constituir o peptídeo maduro de 119 aminoácidos. Outro trabalho importante para a identificação de regiões peculiares na seqüência peptídica do proBDNF foi a descoberta de um motivo não contínuo no domínio maduro ($^{161}\text{I} - ^{18}\text{E} - ^{105}\text{I} - ^{106}\text{D}$) que é necessário para o direcionamento do proBDNF a uma rota regulada. Na ausência deste motivo, o BDNF é direcionado para uma via constitutiva (Lou *et al.*, 2005).

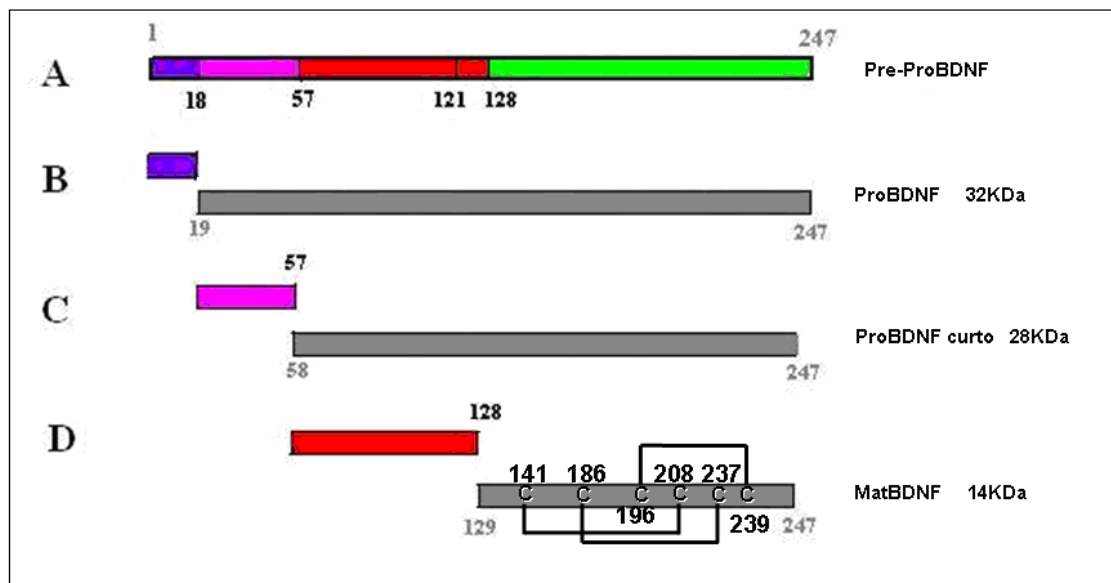


Figura 2. O BDNF é traduzido em uma pré-proteína, constituída de um pré-prodomínio, um pro-domínio e uma forma madura C-terminal. **A)** Pré-proprodomínio com o peptídeo sinal (18 primeiros aminoácidos) para direcionamento ao retículo endoplasmático, mostrando as posições relativas às clivagens (57 e 128) e um sítio para glicosilação na posição 121. **B)** Pro-domínio com características de uma proteína secretória de 229 aminoácidos de comprimento. **C)** proBDNF curto com 190 aminoácidos de comprimento, gerada no retículo endoplasmático. A clivagem pelo pró-hormônio convertase S1P/SKI-1 ocorre na posição ^{57}T , no motivo consenso $^{54}\text{RGLT}\downarrow\text{SL}^{59}$. **D)** Forma madura com 119 aminoácidos de comprimento. A clivagem pela plasmina ocorre na posição ^{128}R , no motivo consenso $^{125}\text{RVRR}\downarrow^{128}$. Esquematizada na figura, está à localização aos pares dos seis resíduos de cisteína que formam as três pontes dissulfeto.

A utilização do proNGF recombinante para elucidar a relação da sua porção pro-peptídica com o domínio maduro foram importantes a fim de obter uma co-relação com o proBDNF. Neste sentido, *Rattenholl e colaboradores*, em 2001, demonstraram o efeito da pro-seqüência do NGF na formação das pontes dissulfeto e no enovelamento do domínio maduro *in vitro* (*Rattenholl et al., 2001a*). Em relação à formação estrutural, análises físico-químicas revelaram que a região pro do NGF forma um dímero, assim como o domínio maduro e, indicaram uma quantidade limitante de estrutura secundária na porção pro (*Rattenholl et al., 2001b*). *Kliemannel e colaboradores* (2004), em pesquisa subsequente, sugeriram que a região pro do NGF parece estruturar-se quando se liga a parte madura, adotando uma estrutura terciária definida. Considerando o enfoque nas características hidrofóbicas, identificaram-se os aminoácidos necessários para associação intramolecular dos domínios pro e maduro do NGF (W⁸³ - A⁶³ na região pro; W¹²¹ na região madura) e, relataram-se propriedades estruturais desta ligação baseada principalmente na superfície exposta ao solvente (*Kliemannel et al, 2007*). *Paiardini e colaboradores* (2008), mensurando esforços para obter uma delimitação estrutural para a porção pro-peptídica, indicaram que a porção N-terminal do proNGF e do proBDNF, são, na maior parte, regiões desordenadas intrinsecamente (IDRs). Com base nesta inferência, da presença de região com padrão de desordem, *Paoletti e colaboradores, em 2009*, obtiveram um modelo do proNGF em camundongo, através da metodologia de SAXS¹, e coletaram informações em relação à estrutura tridimensional (3D), e à dinâmica do proNGF, fornecendo relações entre a estrutura molecular, movimento em nível atômico, e sua função.

¹ SAXS (*Small-Angle X-ray Scattering*), método estrutural aplicável a partículas nativas em solução; é adequado, particularmente, ao estudo de sistemas menos estruturados. Metodologia comumente usada para pesquisar processos, como por exemplo, relacionados ao enovelamento/não-enovelamento de proteínas, ou rotas não totalmente elucidadas (*Bernadó et al., 2007*).

Em relação ao processamento, dois importantes estudos foram relatados: no primeiro, *Nomoto e colaboradores, (2007)*, indicaram que especialmente a parte C-terminal do proNGF e do proBDNF afeta muito a constituição do domínio maduro, por estar envolvida no processamento das neurotrofinas; com o segundo estudo, demonstrou-se que a endocitose, assim como a clivagem são pré-requisitos para a produção da forma processada, capaz de induzir a ativação do receptor Trk, e inferiu-se que sob condições fisiológicas, as pro-neurotrofinas não se ligam ou ativam diretamente tais receptores (*Boutillier et al, 2008*).

A estrutura protéica madura do BDNF contém seis resíduos de cisteína, que formam três pontes dissulfeto (*Robinson et al., 1995*). Os referidos resíduos de cisteína, na qual participam da estruturação característica desta proteína, são extremamente conservados em todas as neurotrofinas. O matBDNF apresenta uma topologia na forma de dímero, com uma interfase dimérica formada por interações hidrofóbicas conservadas, identificadas pela comparação em estudo de superposição (*Robinson et al., 1995; Butte et al., 1998; Robinson et al., 1999*). A estrutura secundária de um monômero do matBDNF é formada por sete *folhas-β*, das quais quatro, formam dois pares de *folhas-β* antiparalelas retorcidas, dispostas longitudinalmente. Estes dois pares de *folhas-β* antiparalelas estão unidos por três loops do tipo grampo (P-hairpin) (loop 1, loop 2, e loop 4), e um loop longo (loop 3) (*Figura 3*).

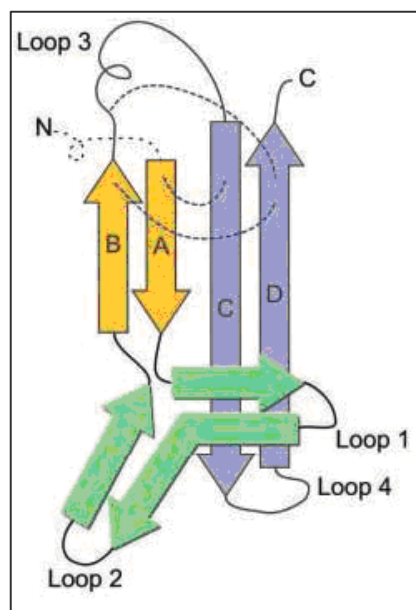


Figura 3. Topologia de um monômero do matBDNF. O esquema mostra os quatro loops, assim como, os dois pares de folhas- β antiparalelas (identificados pelas letras A à D). N e C terminais estão indicados. Os elementos de estrutura secundária típico desta família são constituídos de sete folhas- β . **Fonte:** Butte, 2001.

A porção madura foi o foco de trabalhos iniciais e, inicialmente não se dispunha de informações sobre o pró-domínio, em comparação a atual gama de estudos investigando a importância da região pro no direcionamento para uma rota distinta da forma madura, assim como no auxílio do enovelamento adequado da forma madura. A esta prévia falta de suporte para as pesquisas com o pro-BDNF intacto, corrobora a atual situação que apenas o domínio maduro (região C-terminal do pro-peptídeo), e uma região do gene (referente ao promotor III) foram submetidos à técnica de cristalização de raios-X, que traduz as prováveis configurações dos átomos no cristal (*Tabela 1*).

Em relação ao domínio maduro, houve a publicação em bancos de dados protéicos de complexos cristalográficos heterodiméricos do BDNF com outras duas neurotrofinas: BDNF/NT-3 (*Robinson et al., 1995*), e BDNF/NT-4 (*Robinson et al., 1999*). Estes estudos experimentais identificaram os resíduos do BDNF necessários à ligação aos receptores TrkB (também conhecido como NTRK2) e p75^{NTR}. Os estudos com os

complexos de heterodímeros comprovaram sua existência *in vitro*, mas a atividade biológica *in vivo* ainda não está estabelecida (Robinson et al., 1995; Butte et al., 1998; Robinson et al., 1999). Esses estudos corroboram a idéia que populações neurais específicas sofram a ação do BDNF complexado com outras neurotrofinas para reforçar ou intensificar a sinalização em uma região. Em relação à preferência de formação dos complexos diméricos *in vivo*, destaca-se um estudo que conduziu um ensaio da atividade biológica, em neurônios; este estudo concluiu que houve maior atividade se considerado o BDNF como homodímero (duas unidades simétricas) (Junkbluth et al., 2005). Entretanto, no contexto estrutural, ainda não existe um modelo cristalográfico que englobe dois monômeros do pro-BDNF intacto, assim como do domínio maduro, ou simplesmente da porção pro.

Tabela 1. Estruturas cristalográficas disponíveis com uma porção do gene BDNF, e a porção madura da proteína (matBDNF - resíduos 8-116).

Código de Acesso	Descrição	Resolução da estrutura (Å)	Referência
3C2I	Domínio de ligação da molécula MeCP2 em complexo com uma seqüência metilada do <i>bdnf</i> ²	2.50	Ho et al. (2008)
1B8M / 1HCF	Heterodímero do matBDNF com a NT-4.	2.75	Robinson et al., (1999)
1BND	Heterodímero do matBDNF com a NT-3.	2.30	Robinson et al., (1995)

Levando em consideração todos os estudos com o domínio maduro, e a região pro do NGF, e também do BDNF, ainda se discute algumas questões relevantes. O mecanismo pela qual o pro-BDNF é processado na via intracelular é desconhecido até o momento, mas tendo o conhecimento da atividade biológica distinta das formas secretadas processada e não processada, sugere-se que seja um processo regulado (Gray & Ellis, 2008). Em relação à estabilidade dimérica do pro-peptídeo, há uma hipótese

² Seqüência do *bdnf* compreendendo os nucleotídeos 1 a 40:
Região 5' TCTGGAACGGAATTCTTCTAATAGAAGAATTCCGTTCCAG 3'.

que esta região adota uma estrutura terciária estável dimérica, por estabilizar o padrão dissulfeto em ambos os monômeros do domínio maduro; e este enovelamento seria considerado transitório quando envolvesse somente a formação de pontes dissulfeto em um monômero (*Kliemann et al, 2004*).

1.4. Sinalização mediada pela proteína BDNF.

Nos últimos anos, houve maior interesse em relação ao fato que o proBDNF e o matBDNF são direcionadas para rotas celulares diferentes, através da interação a receptores distintos. O proBDNF possui maior afinidade ao p75^{NTR}, e juntamente com o co-receptor sortilina, favorece uma rota pró-apoptótica (*Teng et al., 2005*). A ligação do matBDNF com o receptor *TrkB* é responsável pelo controle do desenvolvimento e a manutenção dos seus alvos-celulares, refletindo uma sinalização anti-apoptótica (*Reichardt, 2006*). O matBDNF é preferencialmente direcionado para uma rota regulada, e o proBDNF secretado sinaliza para uma via constitutiva (*Labmann & Brigadski, 2009*). A *Figura 4* mostra a interação do BDNF com algumas proteínas, e o tipo de evidência para estas associações, se de cunho experimental ou não.

Importantes estudos, porém com resultados conflitantes, mostraram uma possível relação para o proBDNF e o matBDNF, com relevantes implicações para o direcionamento de rotas. Assim, quando foi estudado a detecção do proBDNF e matBDNF no hipocampo de camundongo sugeriu-se que, o proBDNF é um intermediário transitório, na qual é rapidamente convertido a matBDNF (*Matsumoto et al., 2008*). Já *Yang e colaboradores (2009)*, utilizando a quantificação do BDNF em neurônios de camundongo, sugeriram que o proBDNF não é um intermediário na rota biosintética; a eficiência na

clivagem do proBDNF é regulada ao longo do desenvolvimento, e sua ação pode ser mais forte durante o período pós-natal.

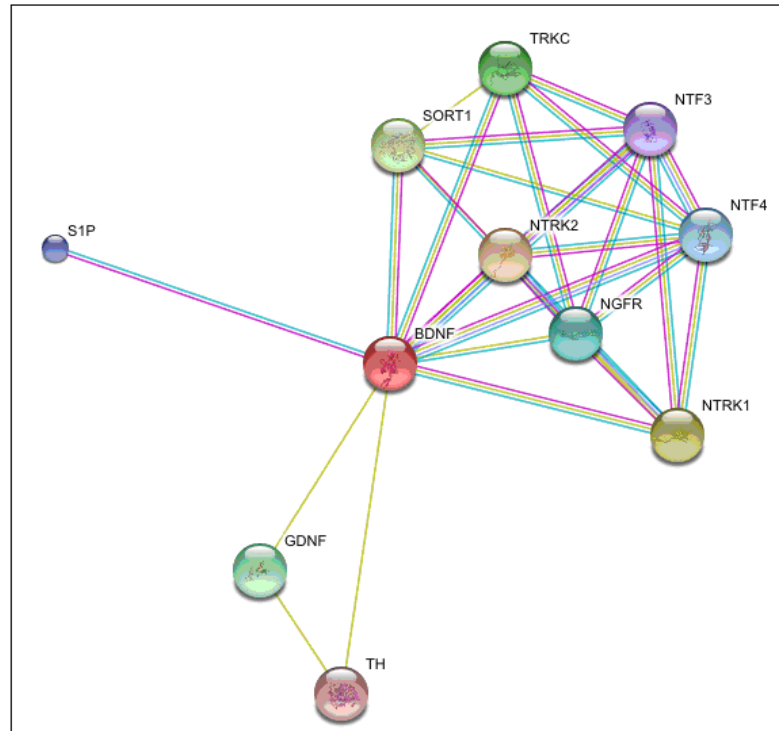


Figura 4. A figura mostra um esquema da interação do BDNF com outras proteínas. **Legenda:** TH: Tirosina 3-Monooxigenase; GDNF: fator neurotrófico derivado da glia; S1P: Precursor da Protease serínica sítio 1; NTRK1: Precursor do receptor do NGF de alta afinidade (também conhecido como TRKA); NGFR: Receptor do NGF; NTRK2: precursor do receptor dos fatores BDNF e NT4; SORT1: Precursor da sortilina (também conhecido como NTR3 – receptor da neurotensina tipo 3); TRKC: precursor do Receptor do fator NT3; NTF3: Neurotrofina tipo 3; NTF4: Neurotrofina tipo 4 (também conhecido como NTF5 – Neurotrofina tipo 5). A cor das linhas representa o tipo de evidência para a associação: Experimentos (em rosa), informações em bancos de dados (em azul), mineração de dados (verde-musgo) e, homologia (roxo claro). **Fonte:** Webserver SMART. Disponível em: www.ebi.ac.uk/embl.

Já vem de longo conhecimento que este peptídeo constitui o principal regulador da transmissão sináptica e plasticidade em várias regiões do SNC. A versatilidade do matBDNF é enfatizada por sua contribuição a uma faixa de respostas adaptativas neurais, incluindo LTP, certas formas de plasticidade de curto prazo, bem como na regulação homeostática da excitabilidade intrínseca neural (Kovalchuk et al., 2002; Pang et al., 2004). Já, o oposto ocorre com o proBDNF, que quando administrado exogenamente, induziu a apoptose em neurônios da periferia (Teng et al., 2005), e facilitou a depressão

de longo-termo (LTD) no hipocampo (*Woo, et al., 2005*). Estudos mais recentes e consistentes sobre o proBDNF mostraram que, a estimulação neural de baixa frequência induziu, predominantemente, a secreção do proBDNF em neurônios do hipocampo, e o aumento extracelular do matBDNF acontecia com a estimulação de alta frequência (*Negappan, et al., 2009*). Dois SNPs raros, resistentes a clivagem para a forma madura – R125M e R127L – mostraram que, o proBDNF exerceu funções distintas, induzindo a apoptose, nos neurônios cerebelares, e reduzindo o número de fibras colinérgicas, nos neurônios colinérgicos basais e hipocampais (*Hoshimizu, H., et al, 2009*).

A função do BDNF foi relatada no envolvimento de algumas doenças neurológicas incluindo a doença de Alzheimer (*Ferrer et al., 1999*), esquizofrenia (*Pillai, 2008*), Parkinson (*Howells et al., 2000*) e Epilepsia (*Binder et al., 2001*). Outros trabalhos demonstraram o envolvimento do BDNF em distúrbios do humor (*Hong et al., 2003; Jiang et al., 2005*), e também no metabolismo diferenciado na dependência de tóxicos (*Bolamos & Nestter, 2004*). Um SNP funcional e amplamente relatado na literatura, comum na seqüência do BDNF, é uma substituição da valina (Val) pela metionina (Met) no códon 66 (Val66Met) na pro-proteína. Esta mutação foi primeiramente relacionada à redução na anatomia do cérebro e à memória episódica pobre (*Egan et al., 2003*), e posteriormente, um estudo com variante BDNF Met em ratos, relacionou-a também à fenótipos de comportamentos depressivos e ansiedade (*Chen et al., 2006*). Ambos os estudos constataram que essa variante foi expressa em níveis normais no cérebro, mas sua secreção para os neurônios foi diminuída. Em relação a associação baseada em análise familiar, variações no *BDNF* – Val66Met e -270C/T - foram associadas à susceptibilidade a anorexia e bulimia nervosa, mostrando um efeito no balanço energético (*Mercader, et al., 2007*).

1.5. Alinhamento de seqüências e caracterização molecular evolutiva.

O alinhamento de seqüências nucleotídicas e protéicas, com graus de identidade, homologia e similaridade diversos, é uma ferramenta que compara duas ou mais seqüências por meio de buscas por uma série de caracteres ou padrões de caracteres que estão na mesma ordem. O alinhamento é utilizado para diferentes tipos de abordagens, tais como, obtenção de padrões de conservação e variação, estudos de comparação entre diferentes espécies, busca pela correta resolução taxonômica, assim como busca por motivos seqüenciais que envolvam aspectos funcionais. O alinhamento pode ser constituído de seqüências de um mesmo segmento gênico ou de seu produto, ou até mesmo incluir vários genes ou proteínas. Os recursos computacionais que realizam os alinhamentos permitem aperfeiçoar os dados, através do tratamento com gaps (representados pelo símbolo -), através da obtenção de *scores* para avaliar o resultado do alinhamento, e também, através da obtenção de valores de probabilidade atribuídos na matriz de substituição de blocos de aminoácidos conservados conhecida como matriz Blosum (*Lesk, 2008*). Esta última é o método mais indicado para alinhamentos locais, e possui a vantagem de ser isenta de extrapolações. As seqüências sujeitas ao alinhamento são encontradas catalogadas em bancos de dados, e a busca pelas referidas seqüências é feita utilizando algoritmos, como por exemplo, BLAST (*Ewens & Grant, 2001*). Este método permite comparar a seqüência-alvo com uma biblioteca ou um banco de dados de seqüências baseado na similaridade entre elas, e se utiliza do alinhamento local (localiza fragmentos de seqüências que são mais similares). BLAST é o programa mais usado em bioinformática, e tem a vantagem de ser rápido;

Uma porção do gene *BDNF* tem sido usada como marcador filogenético em diversos estudos de mamíferos, o que levou a um acúmulo de dados sobre estes locos. Como consequência houve uma crescente gama de publicações de seqüências do *BDNF* em diversas espécies, que não foram objetos de análises detalhadas. Atualmente, a filogenia de algumas famílias ou espécies dentro do grande grupo não está bem resolvida. Sendo assim, as informações da correta classificação taxonômica são essenciais para futuros estudos de genômica comparativa entre as famílias, e para auxiliar futuros estudos das espécies não incluídas em abordagem molecular.

Tendo em vista estas considerações, *Kullander e colaboradores*, em 1997, conduziram um estudo com seqüências gênicas do *BDNF* (256 bp) pertencentes a espécies de quatro ordens da infraclasse dos marsupiais (Metatheria), estabelecendo a correta classificação entre elas. Juntamente com outras análises utilizando os outros membros das neurotrofinas e, o estudo da ordem Monotremata (Prototheria), foi demonstrado que estas seqüências são bem adequadas para análises filogenéticas dos mamíferos, podendo resolver a relação basal entre eles (*Kullander et al., 1997*). Já em 2001, *Murphy e colaboradores* resolveram boa parte da relação inter-ordem dos placentários (Eutheria), analisando uma grande demanda de seqüências nucleares, dentre elas o *BDNF*, a qual tem colaborado para estudos de interpretação do padrão biogeográfico e processos evolucionários, envolvidos na grande gama de diversificação desta infraclasse (*Murphy et al., 2001*). Dentro dos mamíferos placentários, os morcegos (ordem Quiróptera) perfazem mais que 20% dos mamíferos existentes (*Miller-Butterworth, et al, 2007*); a taxonomia e a filogenia do gênero *Miniopterus* foram abordadas em estudos incluindo caracteres morfológicos, de sítios de restrição de rRNA, de mDNA, e de íntrons nucleares, na qual o posicionaram diferentemente em relação a famílias de morcegos (*Kawai et al. 2002; Van Den Bussche & Hooper, 2004; Eick et al. 2005*). Após

um crescente esforço, através da análise das seqüências de representantes das famílias de morcegos, com base em 16 genes nucleares, entre eles o *BDNF*, houve a confirmação que o gênero *Miniopterus* pertence à família Miniopteridae (Miller-Butterworth, et al, 2007). Outro grupo representativo dos mamíferos inclui os Mustelídeos, pertencente à família mais rica em espécies dentro da ordem carnívora e, compreende 59 espécies classificado em 22 gêneros, também abordado nesta pesquisa; a extensa diversidade e característica evolucionária de radiação adaptativa têm contribuído para a dificuldade na resolução de sua história filogenética (Givnish & Sytsma, 1997). Sendo assim, utilizando-se da análise de 22 segmentos gênicos, entre eles o *BDNF*, os mustelídeos foram consistentemente classificados dentro de quatro maiores clados e três linhagens monotípicas (Koepfli, et al., 2008). Através da análise desse quadro de classificação houve a inferência que esta família sofreu dois períodos de diversificação coincidindo com a maior mudança paleo-ambiental e biótica que ocorreram durante o Neogêneo (Koepfli, et al., 2008).

Estes foram alguns exemplos de grupos avaliados nas análises de padrões de nucleotídeos e de aminoácidos nesta pesquisa, sendo que trabalhos na qual abordaram as questões de correta taxonomia e filogenia descritos acima são complementares para as análises comparativas. Os padrões das seqüências alinhadas podem ser analisados através de diferentes enfoques, como por exemplo, a proporção de códons conservados e variados, e a indicação do tipo de seleção natural atuante nos diferentes códons. A seleção negativa implica graus diferenciados de restrição a mudanças, e na seleção positiva ocorrem taxas de mudanças que seriam maiores que as encontradas ao acaso, permitindo variabilidade e adaptação. Ambos os tipos de seleção apresentam relevância funcional, sendo que o tipo mais comum de ser detectado ao longo do DNA é aquela que em que a proteína não perde a identidade funcional (seleção negativa). O padrão de seleção natural mapeado é obtido com base em modelos estatísticos de substituição de

códons, que diferem no número de substituições não-sinônimas por códon não sinônimo sobre o número de substituições sinônimas por códon sinônimo (dN/dS) conhecido como estimativa de ômega (ω). O modelo M0 assume o mesmo valor de ômega para todos os códons; M1 assume um valor de ω igual a 0, indicando códons conservados, e também outro valor de ω igual a 1, que indica códons neutros; M2 assume $\omega > 1$, inferindo a ocorrência de seleção natural positiva; M3, M7 e M8 selecionam códons conservados, neutros e, positivos. Estes modelos também diferem quanto à distribuição em relação à classe de códons, podendo ter distribuição normal (modelos M1 e M2), distribuição discreta (M0 e M3), e distribuição beta (M7 e M8). Os modelos aos pares são usados para calcular o teste de razão de verossimilhança (Likelihood Ratio Test – LRT) e indicar qual ou quais dos modelos melhor representa o conjunto de dados amostrados.

1. 6. Modelagem molecular por homologia.

Algumas proteínas-moldes, por possuírem considerável grau de similaridade na seqüência primária de aminoácidos, e por adotarem um padrão de conformação estrutural conservado (*Nayeem et al., 2006*) são utilizadas para modelar proteínas-alvo, na qual se quer determinar a estrutura tridimensional. Este procedimento para a construção de modelos estruturais é conhecido como modelagem molecular por homologia ou modelagem comparativa (*Martí-Renom et al., 2000; Deane & Blundell, 2003*). A seqüência protéica alvo é submetida a um alinhamento com proteínas que já dispõem de estruturas tridimensionais definidas, disponíveis no banco de dados de proteínas PDB (Protein Data Bank) (*Berman et al., 2000*), a fim de escolher a(s) estruturas tridimensionais - template(s) - mais apropriada(s) aos objetivos de uma pesquisa. Estes três primeiros procedimentos: identificação, seleção, e alinhamento da seqüência-alvo com o(s) template(s) podem ser

conduzidos através de uma busca no banco de dados protéico online do *National Center of Biological Information* (NCBI) (<http://blast.ncbi.nlm.nih.gov>), utilizando o algoritmo psi-blast para a procura por seqüências homólogas, e a matriz de substituição bluson45 para o alinhamento de seqüências de proteínas. Outros softwares mais específicos que o BLAST são então utilizados para a obtenção de um alinhamento ótimo, como o programa online T-coffee (www.ebi.ac.uk/t-coffee).

O template escolhido(s) é então submetido a técnicas de otimização de distâncias geométricas a fim de satisfazer as restrições espaciais baseado no alinhamento molde/alvo. Este método de construção de modelos é implantado em softwares como, por exemplo, Modeller, na qual é o mais utilizado para pesquisas em modelagem (*Sali & Blundell, 1993*). A determinação das regiões estruturalmente conservadas das proteínas-templates e da proteína recém modelada é realizada através de métodos de ajuste dos quadrados mínimos (*"least-squares fitting methods"*) pela sobreposição de uma proteína sobre a outra (*Da Silva & Silva, 2007*).

As etapas de escolha de um molde, alinhamento e construção do modelo devem ser repetidas, quantas vezes sejam necessárias para a construção de um modelo satisfatório (*Martí-Renom et al., 2000*). Possíveis correções na geometria e na estrutura modelada podem ser detectadas através de campos de força empíricos, e corrigidas por métodos como minimização de energia ou dinâmica molecular. Estes parâmetros são implementados, por exemplo, no programa SANDER do pacote AMBER 9.0 (*Pearlman et al., 1995*). Após este processo de otimização do modelo torna-se importante a subsequente avaliação dos níveis de qualidade e confiabilidade dos modelos gerados, tendo como base para os cálculos, estruturas tridimensionais resolvidas (*Da Silva & Silva, 2007*). Em relação à qualidade, é avaliada a exatidão de parâmetros estereoquímicos, como por exemplo, comprimento das ligações, ângulos entre as ligações, ângulos

torsionais, e quiralidade dos aminoácidos (*Höltje, et al., 2003*); estes parâmetros podem ser mensurados utilizando o programa Procheck (*Laskowski et al., 1993*). Já a confiabilidade dos modelos gerados pode ser avaliada através dos softwares VERIFY3D (*Luethy et al., 1992*), Errat (*Colovos & Yeates, 1993*), e, visualizados através dos softwares Swiss-PdbViwer (*Guex & Peitsch, 1997; Schwede, et al, 2003*), e PyMOL (*DeLano, 2002*).

Finalmente, esta revisão bibliográfica acerca de pontos relevantes sobre o BDNF ajudou o direcionamento desta pesquisa para uma caracterização e elucidação de aspectos desta sequência. Estudos conjuntos enfocando as análises comparativas e estruturais, e suas implicações com características funcionais de uma proteína, podem contribuir para elucidar aspectos desconhecidos desta molécula. Neste contexto, análises comparativas das sequências codificantes do BDNF, podem ajudar a compreender em maiores detalhes aspectos evolutivos, assim como, redirecionar os estudos com esta sequência. Nesta pesquisa, foi conduzida uma caracterização comparativa, focando em regiões conservadas e variadas, e padrões de seleção natural (positiva, negativa, ou neutra). Esta análise foi relacionada a uma caracterização estrutural paralela para abordar a conformação estrutural das diferentes porções do BDNF (pro e maduro).

1.7. Objetivos.

2.1 Objetivos Gerais

2.1.1. Caracterizar a diversidade molecular, padrões de conservação e história evolutiva do gene *BDNF* em mamíferos através da comparação de seqüências representando múltiplas espécies e diferentes linhagens deste grupo.

2.1.2. Caracterizar aspectos estruturais da proteína BDNF de mamíferos utilizando simulações computacionais, e relacionando-os aos padrões observados nas análises evolutivas.

2.2 Objetivos Específicos

2.2.1. Investigar padrões de variabilidade em nível de nucleotídeos e aminoácidos ao longo da região codificadora do gene BDNF em ≥ 100 espécies de mamíferos representando diferentes linhagens.

2.2.2. A partir destas análises comparativas, identificar e caracterizar regiões com altos níveis de conservação ou variabilidade acentuada, e investigar seu potencial envolvimento em aspectos funcionais das formas protéicas resultantes através de análises estruturais.

2.2.3. Caracterizar a estrutura 3D do pró-domínio do BDNF através de um estudo de modelagem molecular estática.

Capítulo 2 – Artigo Científico

ProBDNF matters: Evolutionary and structural bioinformatics provide novel insights
on mammalian Brain-Derived Neurotrophic Factor

Fabíola S. Mattei, Osmar N. Souza, Eduardo Eizirik

Submetido ao periódico **Nature Neuroscience**

ProBDNF matters: Evolutionary and structural bioinformatics provide novel insights on mammalian brain-derived neurotrophic factor

Fabíola S. Mattei^{1,2}, Osmar Norberto de Souza², Eduardo Eizirik¹

¹ Laboratório de Biologia Genômica e Molecular, Faculdade de Biociências, PUCRS. Porto Alegre, RS 90619-900, Brazil.

² Laboratório de Bioinformática, Faculdade de Informática, PUCRS. Porto Alegre, RS 90619-900, Brazil.

Abstract

Multiple studies have addressed the function of brain-derived neurotrophic factor, yet most have only included data from humans or model organisms. Here we apply bioinformatics approaches to model the structure of human proBDNF and to analyze the evolution of this protein in mammals. We show that strong selective constraint occurs on both the pro and mature domains of the protein, and find evidence of adaptive evolutionary change in the former.

Main text

A great many recent studies have focused on the structure and function of the brain-derived-neurotrophic factor (BDNF), including its implication in human disorders and analyses of its developmental and regulatory roles in the nervous systems of model animals¹⁻⁵. Although this molecule has been clearly implicated in many important functions, several aspects of its biological roles remain poorly understood. In particular, the precise role of proBDNF (the longer, pre-cleavage form of the protein) is still contentious⁶⁻⁸, with limited data available on this molecule relative to the mature form of the peptide (matBDNF). In addition, very little is still known with respect to the structure and function of BDNF in non-model species, so that a potential wealth of comparative information on this important molecule remains virtually untapped.

Comparative molecular data may clarify patterns of conservation at the DNA or protein levels, indicating regions that may be under strong evolutionary constraint (i.e. negative selection), implying functional relevance, as well as others that bear signatures of adaptive change (i.e. positive selection)⁹. In this context, a powerful approach to understand biological function of target molecules is to combine methods that detect the strength and pattern of natural selection at each codon with three-dimensional models of protein structure associated with experimentally validated information on active motifs. Such a combined approach has not yet been applied to BDNF, despite the interest in characterizing in detail its structure and function. Moreover, although various structural aspects of BDNF have been characterized (e.g. refs. 10,11), no current model of its structure includes the pro domain, hampering assessments of the biological impact of protein variants affecting this region of the molecule.

In parallel with the efforts to characterize BDNF structure and function, partial or complete DNA sequences from the gene encoding this protein have been used in many phylogenetic studies. The first of those studies, published by our group in 2001, addressed the relationships among placental mammals¹², and led to an intriguing observation regarding the structure of BDNF: two different mammalian lineages (caviomorph rodents such as guinea pigs, and cetartiodactyls such as cows and pigs) exhibited up to 11 extra serine residues (relative to the human sequence) inserted at the same position in the protein (near the cleavage site yielding a shorter proBDNF – see Table S1; Figs. S1, S4), rendering the resulting peptide potentially very different from those of model organisms. This finding led us to investigate in detail the levels of sequence variation in BDNF across mammals, to perform site-by-site analyses of natural selection, and to construct a novel structural model of a complete proBDNF peptide onto which to map the detected patterns of constraint and modification.

We specifically aimed to address five questions: (i) how much selective constraint (i.e. functional relevance) can be detected across proBDNF? (ii) is there any evidence of positive selection at any site? (iii) what is the 3D structural distribution of amino acid sites inferred to be under selection? (iv) are the detected serine expansions located near any functionally relevant site? (v) is there evidence for different evolutionary forces acting upon BDNF in different mammalian groups?

To address these issues we undertook a two-pronged approach, which led us to construct a novel 3D model of homodimeric proBDNF, and to perform multiple comparative analysis of BDNF sequence variation across mammals (see Supplementary

Methods). Briefly, our modeling effort used coordinates from two templates: *2ia8* was selected to model the pro domain, and *1b8m* was used for the mature portion (see Fig. S2; Tables S2-S3). Our models were consistent with SAXS experiments of proNGF¹³. Furthermore, we superimposed the validated mature domain onto different models obtained in modeling (n=10), and confirmed the flexibility of the pro domain, which is best represented as a random coil (Fig. 1; Fig.S3).

The evolutionary analysis of natural selection employed *BDNF* DNA sequences from 153 different mammalian species (Tables S4-S9), which were sorted into six different data sets: (i) data set 1 contained all available complete coding sequences for *BDNF* (n=20); (ii) data set 2 was more inclusive, incorporating complete or partial sequences (spanning ≥ 489 bp) from 94 different species; (iii – vi) specific data sets focused on the orders Carnivora (n= 84), Cetartiodactyla (n= 28), Chiroptera (n= 28), and Rodentia (n= 15), whose current sampling allowed for detailed comparative analyses. Preliminary analysis of sequence variation indicated that both the mature and pro regions of *BDNF* contained a substantial portion of conserved sites at the DNA and protein levels (Table S10). Even though conservation was more extreme in the mature portion of the protein, 68% of the amino acids in the pro domain were completely identical across all major mammalian lineages (data set 1), indicating considerable constraint in this domain as well. We then performed codon-by-codon analyses of natural selection, employing several maximum likelihoods and Bayesian approaches (see Supplementary Methods for details). After selecting the model that provided the best fit to the data (Table S11), we analyzed the ratio of nonsynonymous (amino acid changing) to synonymous (silent) substitutions ($\omega = dN/dS$) for each codon in each data set. Ratios significantly below 1 indicate negative selection (functional constraint), while those significantly above 1 indicate positive selection.

Many codons in both the pro and mature domains were inferred to be under very strong negative selection (Fig. 1; Tables S12-S13). All the amino acid sites that have been experimentally implicated in *BDNF* function (e.g. cleavage sites and the human codon 66, involved in the well-known Val/Met polymorphism^{2,5}) had ratios close to zero, corroborating their strong constraint and validating our approach. Many other sites showed similar levels of constraint, including those adjacent to the observed serine expansions. This indicates that those sites are functionally relevant, and that the insertion of multiple extra residues in that region may lead to biologically meaningful consequences. This suggests that experimental analysis of *BDNF* in mammals bearing those insertions (e.g. guinea pig, cow,

pig) may provide novel insights onto the function and regulation of this molecule. Moreover, the full set of codons observed to be under extreme negative selection may be used to help design experiments that assay their activity directly. Many of these sites are located in the pro domain, supporting the view that this portion has important biological functions.

Another interesting result was the detection of five codons in *BDNF* that may be under positive selection (Figs. 1 and S4; Tables 1, S12, S13), implying that they have undergone adaptive change in at least some of the sampled species. Remarkably, all of these sites were located in the pro domain, indicating that this portion of the protein has been influenced by recent adaptive evolution in different mammalian lineages. Site 84 yielded the most consistent pattern, which was detected in most data sets, particularly in data set two (Table 1), whose sample size provided the most power. Differences among taxon-specific data sets indicate that some sites may have undergone distinct selective regimes in different mammalian lineages, possibly due to contrasting evolutionary pressures affecting the regulation and development of their nervous system. Two interesting examples are codons 51 and 53 (both located near the cleavage site for a short proBDNF): the former bears very strong evidence of positive selection in Cetartiodactyla and negative selection in the other taxa; the latter shows signs of positive selection in rodents, and negative selection in others. Interestingly both of these orders contain serine/threonine expansions in that region of the protein (see Tables S1, S12-S13). These and other results outlined here open up new research avenues allowing further dissection of the biological roles of BDNF in mammals.

References

1. Lee, R., Kermani, P. Teng, K.K., Hempstead, B.L. *Science* **294**, 1945-1948 (2001).
2. Egan, M.F. et al. *Cell* **112**, 257–269 (2003).
3. Lu, B., Pang, P.T., Woo, N.H. *Nature Rev. Neurosci.* **6**, 603-614 (2005).
4. Hu, Y. & Russek, S. *J. Neurochem.* **105**, 1-17 (2008).
5. Soliman, F. et al. *Science.* **327**, 863-865 (2010).
6. Matsumoto, T. et al. *Nature Neurosci.* **11**, 131-133 (2008).
7. Yang, J. et al. *Nature Neurosci.* **12**, 113-115 (2009).
8. Barker, P.A. *Nature Neurosci.* **12**, 105 – 106 (2009).
9. Yang, Z. *Curr. Opin. Genet. Dev.* **12**, 688 – 694 (2002).
10. Lou, H. et al. *Neuron.* **45**, 245 – 255 (2005).
11. Paiardini, A., Caputa, V. *Neuropeptides* **42**, 205 – 214 (2008).
12. Murphy, W. et al. *Nature* **409**, 614 – 618 (2001).
13. Paolletti, F. et al. *Proteins* **75**, 990 – 1009 (2009).

Table 1. Detection of specific amino acid sites in BDNF likely to be under positive selection in at least one of our mammalian data sets. For each of the five sites in each of the data sets, we indicate the probability of belonging to a category with $\omega > 1$, as well as the mean posterior estimate of ω , which were calculated with model M8 and the BEB option in the CODEML program (see Supplementary Methods for details). Site numbers and the respective amino acids follow the observed states in the human sequence (see Tables S12-S13 for a complete list of values for all sites).

Data set	Selected amino acid ^a	Probability of positive selection ^{b,c}	Mean posterior ω (\pm SE) ^b
Data set 1	47 Asn	0.02072	0.332 \pm 0.261
	51 Ala	0.12969	0.535 \pm 0.429
	53 Ser	0.00765	0.124 \pm 0.182
	<u>80 Asn</u>	0.52648	1.092 \pm 0.489
	<u>84 Asn</u>	0.72621	1.320 \pm 0.501
Data set 2	47 Asn	0.00006	0.371 \pm 0.138
	51 Ala	0.00027	0.336 \pm 0.154
	<u>53 Ser</u>	0.77941	1.329 \pm 0.331
	<u>80 Asn</u>	0.59787	1.205 \pm 0.369
	<u>84 Asn</u>	0.98596	1.496 \pm 0.101
Cetartiodactyla	47 Asn	0.00427	0.082 \pm 0.144
	<u>51 Ala</u>	0.95891	2.423 \pm 1.195
	53 Ser	0.01884	0.121 \pm 0.311
	80 Asn	0.09573	0.353 \pm 0.595
	<u>84 Asn</u>	0.81383	2.094 \pm 1.218
Rodentia	47 Asn	0.02816	0.274 \pm 0.296
	51 Ala	0.00476	0.117 \pm 0.164
	<u>53 Ser</u>	0.84950	1.636 \pm 0.736
	80 Asn	0.07289	0.466 \pm 0.362
	<u>84 Asn</u>	0.88530	1.651 \pm 0.629
Chiroptera	<u>47 Asn</u>	0.61496	1.169 \pm 0.464
	51 Ala	0.07566	0.370 \pm 0.391
	53 Ser	0.37821	0.871 \pm 0.535
	80 Asn	0.14180	0.568 \pm 0.432
	<u>84 Asn</u>	0.78233	1.346 \pm 0.399
Carnivora	47 Asn	0.24152	0.662 \pm 0.578
	51 Ala	0.00613	0.086 \pm 0.150
	53 Ser	0.00289	0.077 \pm 0.115
	<u>80 Asn</u>	0.46473	1.038 \pm 0.626
	84 Asn	0.04190	0.219 \pm 0.336

^a Underlined sites indicate amino acids detected as likely being under positive selection in a given data set.

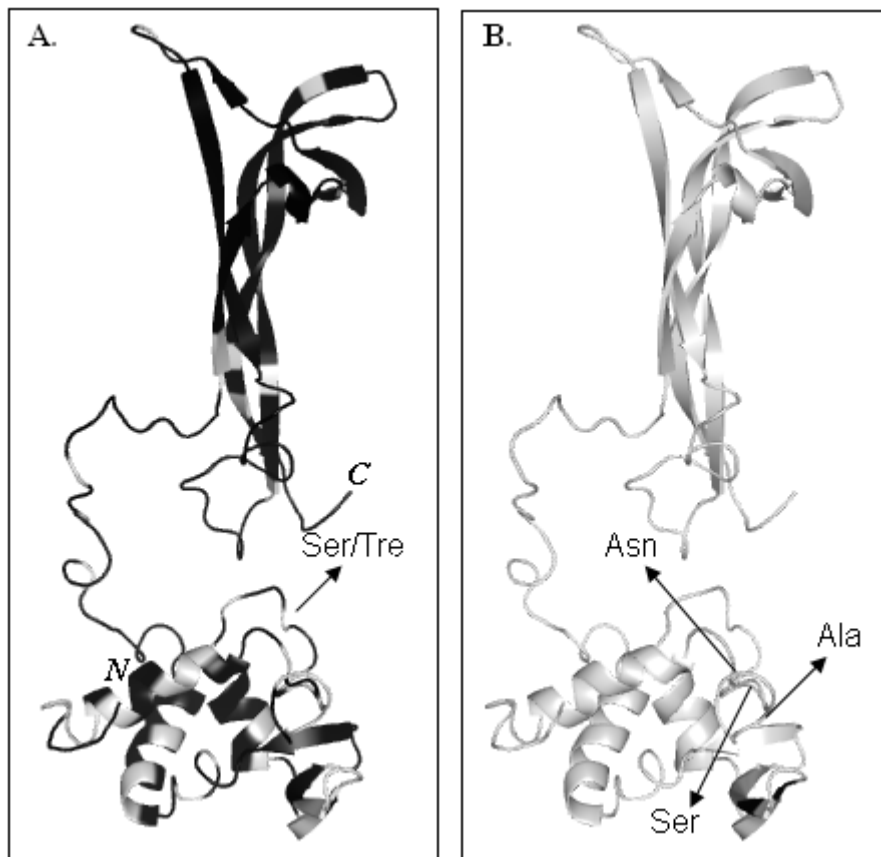
^b Values in bold indicate evidence of positive selection (probability > 0.9, or posterior mean >1).

^c The probability of the site being under negative selection is 1 minus the value indicated here (e.g. 0.97928 for site 47 in data set 1 – see Table S11 for the mean ω of each site category in each data set).

Figure Legend

Figure 1. Three-dimensional structure of a human proBDNF monomer, modeled using pdb templates *1b8m* subunit A and *2ia8*. The top portion of each panel corresponds to the mature domain of BDNF, while the intermediate loop and the bottom portion contain the pro domain. Colors indicate the ω value (*i.e.* the non-synonymous to synonymous ratio of nucleotide substitution) for each codon, calculated using data set 1 and the M8 model (see Supplementary Methods for details). **A.** Codons exhibiting the lowest ω values (0.053 to 0.06), implying the strongest selective constraint, are shown in black; all others are shown in gray. The arrow represents the point where serine/threonine expansions have been observed in two mammalian lineages, exactly where cleavage occurs to form a shorter proBDNF molecule (see Tables S1 and S12). **B.** Codons exhibiting the highest values of ω (1.0 to 1.32), implying evidence of positive selection, are shown in black; all other residues are shown in *gray*. Arrows indicate three other sites (Asn47, Ala51, and Ser53) bearing ω values > 1 in other data sets (see Tables 1, S12 and S13).

Figure 1



Online Supplementary Information – Mattei et al. 2010

Summary for Supplementary Information

1. Supplementary methods	2
2. References for the Supplementary methods	4
3. Supplementary Tables	5
3.1. Table S1. Summary of serine/threonine insertions.....	5
3.2. Tables S2 and S3. Summary of structural model parameters....	6
3.3. Table S4 – S9. Data set composition.....	7
3.3.1. Supplementary references for tables S4 – S9.....	17
3.4. Table S10. Proportion of conserved sites.....	18
3.5. Table S11. Parameters of Likelihood Ratio Tests.....	19
3.6. Tables S12-S13. Full list of posterior probabilities and ω values for all codon sites.....	22
4. Supplementary Figures	34
4.1. Figure S1. Sequence of human BDNF <i>isoform a</i>	34
4.2. Figure S2. Alignment of templates and target-sequence.....	35
4.3. Figure S3. Structure of a proBDNF homodimer	36
4.4 Figure S4. Location of positively selected amino acids.....	37

Supplementary Methods

ProBDNF Homodimer Model Construction and Evaluation

A proBDNF structural model was constructed using a comparative molecular modeling technique¹ supported by the swiss-pdbViewer version 3.72 and Pymol version 0.99rc6³ visualization software packages. We initially generated the mature BDNF homodimer from the crystallographic structure of a BDNF/NT4 heterodimer (*pdb ID: 1b8m*), using information on the mature BDNF monomer (chain A), and applying the copy and translate technique. To create a complete 3D model, we conducted an online screening procedure with psi-blast algorithms (<http://blast.ncbi.nlm.nih.gov>) to search for the best available templates that covered the target-sequence of human proBDNF isoform a (*gi:219842314*), using the protein data bank (PDB), the blosum45 matrix, and threshold 0.5. After selection, the two templates (*pdb ID: 1b8m* subunit A, and *2ia8*) were used for modeling the proBDNF. The optimal alignment was performed with the program T-coffee (www.ebi.ac.uk/t-coffee). ProBDNF models were constructed using the software Modeller4, which builds 3D models by meeting specific constraints in the target-sequence, based on alignments with the template sequences. Finally, all generated models were evaluated using Procheck 5 (Tables S2-S3).

Comparative Sequence Analysis

Mammalian BDNF coding sequences (CDS) were obtained from GenBank entries (<http://www.ncbi.nlm.nih.gov>), using the BLAST algorithm applied to the nucleotide collection database. Sequences from 153 species were retrieved, including complete CDS, partial CDS or full mRNAs. Species names, sequence lengths, GenBank accession numbers and the original reference are presented in Tables S4-S9. These sequences were aligned using the CLUSTALW algorithm⁶, manually inspected, and sorted by sequence length or taxonomic assignment, representing different mammalian lineages. Sequences with ambiguous characters were eliminated from the analysis. All analyses were conducted excluding the signal peptide (first 18 codons). The alignments sorted by length were divided into two groups: i) data set 1, which included species with complete coding sequences; and ii) data set 2, which was selected to represent 24 mammalian lineages (with each sequence spanning at least 489 bp). The alignments sorted by taxon

represented mammalian orders which currently present a reasonable sample of available species: Carnivora, Cetartiodactyla, Chiroptera and Rodentia. Data sets were organized using the package MEGA 4.07, which was also employed to run exploratory analyses of sequence conservation.

Natural Selection Analysis

More detailed molecular evolutionary studies were conducted with the CODEML program within the PAML4 package⁸, which provides log-likelihood (lnL) values for different models of codon evolution and posterior probabilities for distinct selection regimes (negative, neutral, positive) at each site. Specifically, we provided the program with unrooted Neighbor-Joining trees estimated in MEGA with Maximum Composite Likelihood distances, including all codon positions. All positions containing alignment gaps and missing data were eliminated in a pairwise fashion (*pairwise* deletion option).

We conducted Likelihood Ratio Tests (LRT) to assess model fitting. These tests employ as focal statistics twice the difference between the log likelihoods (lnL) of contrasted models: M1 vs. M2, and M7 vs. M8 (only nested models should be compared with this approach), with a χ^2 null distribution and degrees of freedom equal to the difference in the number of parameters⁹. The M7-M8 comparison is a stringent test of positive selection, compared with M1-M2. The M0-M3 comparison, however, is more a test of variable selective pressure among sites than a test of positive selection¹⁰.

The omega (ω) parameter was estimated using the option NSsites. This specifies models that allow the dN/dS ratio (ω) to vary among sites^{11,12}. We analyzed six NSsites models which differed with respect to proportion of site classes (p) and their respective mean ω value. Models were as follows: M0 (one-ratio) assumes the same ω ratio for all sites; M1 (Nearly Neutral) assumes $\omega=0$ (conserved sites) and $\omega=1$ (neutral sites); M2 (Positive selection) allows conserved sites, neutral sites and positive sites, the latter with $\omega>1$ (independent); M3 (Discrete) assumes these same site classes, but following a discrete distribution, with both the proportion of classes and ω parameters estimated from the data; M7 (beta) and M8 (beta& $\omega>1$) assume a beta distribution of ω values, with M8 allowing the presence of positively selected sites. An empirical Bayes (EB) approach was then used to calculate the posterior probability of each codon belonging to a specific site class.

References for the supplementary methods

1. Martí-Renom, M., Stuart, A., Fiser, A. et al.. *Annu. Rev. Biophys. Biomol. Struct.*, 29, 291 – 325 (2000).
2. Guex, N., Peitsch, M. *Electrophoresis*, 18, 2714 – 23, 1997.
3. DeLano, W.L. *The PyMOL Molecular Graphics System* (2002) DeLano Scientific, San Carlos, CA, USA.
4. Sali, A., Blundell, T. J. *Mol. Biol.* 234, 779 – 815 (1993).
5. Laskowski R A, MacArthur M.W, Moss D. S, Thornton J. M. *J. App. Cryst.* 26, 283 – 291 (1993).
6. Thompson, J., Higgins, D., Gibson, T. *Nucleic Acids Research.* 22, 4673 – 80 (1994).
7. Tamura K.; Dudley J.; Nei, M; Kumar, S. *Mol Biol Evol.* 24, 1596 – 99 (2007).
8. Yang, Z. *Mol Biol Evol.* 24, 1586 -91 (2007).
9. Yang, Z. *Mol. Biol. Evol.* 15, 568 – 73 (1998).
10. Anisimova, M.; Bielawski, J.; Yang, Z. *Mol Biol Evol.* 18, 1585 – 92 (2001).
11. Nielsen, R.; Yang, Z. *Genetics.* 148, 929 – 936 (1998).
12. Yang, Z., Nielsen, R, Goldman, N, Pedersen, A. *Genetics.* 155, 431 – 449 (2000).

Supplementary Tables.

Table S1. Serine and/or threonine insertions in BDNF sequences of selected mammalian species. All sequences have amino acid position 58 (numbering according to the human sequence) as the starting point, after which some species exhibit inserted sequences coding for multiple extra residues.

Species	Taxon	S / T insertion
<i>Balaena mysticetus</i>	Cetartiodactyla	S ⁵⁸ S
<i>Eubalaena japonica</i>	Cetartiodactyla	S ⁵⁸ S
<i>Balaenoptera acutorostrata</i>	Cetartiodactyla	S ⁵⁸ S
<i>Balaenoptera borealis</i>	Cetartiodactyla	S ⁵⁸ S
<i>Balaenoptera edeni</i>	Cetartiodactyla	S ⁵⁸ S
<i>Balaenoptera musculus</i>	Cetartiodactyla	S ⁵⁸ S
<i>Balaenopeta physalus</i>	Cetartiodactyla	S ⁵⁸ S
<i>Eschichtius robustus</i>	Cetartiodactyla	S ⁵⁸ S
<i>Megaptera novaeangliae</i>	Cetartiodactyla	S ⁵⁸ S
<i>Caperea marginata</i>	Cetartiodactyla	S ⁵⁸ S
<i>Delphinus tropicalis</i>	Cetartiodactyla	S ⁵⁸ S
<i>Souza chinensis</i>	Cetartiodactyla	S ⁵⁸ S
<i>Stenella coeruleoalba</i>	Cetartiodactyla	S ⁵⁸ S
<i>Mesoplodon densirostris</i>	Cetartiodactyla	S ⁵⁸ S
<i>Lipotes vexillifer</i>	Cetartiodactyla	S ⁵⁸ S
<i>Neophocaena phocaenoides</i>	Cetartiodactyla	S ⁵⁸ S
<i>Physeter catodon</i>	Cetartiodactyla	S ⁵⁸ S
<i>Pontoporia blainvillei</i>	Cetartiodactyla	S ⁵⁸ S
<i>Tursiops aduncus</i>	Cetartiodactyla	S ⁵⁸ S
<i>Tursiops truncatus</i>	Cetartiodactyla	S ⁵⁸ S
<i>Sus scrofa</i>	Cetartiodactyla	S ⁵⁸ SSSSS
<i>Okapia johnstoni</i>	Cetartiodactyla	S ⁵⁸ SSS
<i>Bos taurus</i>	Cetartiodactyla	S ⁵⁸ SSS
<i>Tragelaphus eurycerus</i>	Cetartiodactyla	S ⁵⁸ SSS
<i>Hippopotamus amphibius</i>	Cetartiodactyla	S ⁵⁸ S
<i>Rhinolophus creaghi</i>	Chiroptera	S ⁵⁸ S
<i>Hipposideros commersoni</i>	Chiroptera	S ⁵⁸ S
<i>Macroderma gigas</i>	Chiroptera	S ⁵⁸ S
<i>Macroderma lyra</i>	Chiroptera	S ⁵⁸ S
<i>Craseonycteris thonglongyai</i>	Chiroptera	S ⁵⁸ S
<i>Rhinopoma hardwickei</i>	Chiroptera	S ⁵⁸ S
<i>Cavia porcellus</i>	Rodentia	S ⁵⁸ SSSSSSSS
<i>Cavia tschudii</i>	Rodentia	S ⁵⁸ SSSSSSSS
<i>Agouti taczanowskii</i>	Rodentia	S ⁵⁸ SSSSSSSSSSSS
<i>Hydrochaeris hydrochaeris</i>	Rodentia	S ⁵⁸ SSSSSSSS
<i>Hystrix brachyurus</i>	Rodentia	S ⁵⁸ SSSSS
<i>Erethizon dorsatum</i>	Rodentia	S ⁵⁸ SS
<i>Dinomys branickii</i>	Rodentia	S ⁵⁸ SS
<i>Mus musculus</i>	Rodentia	T ⁵⁸ TS
<i>Rattus norvegicus</i>	Rodentia	T ⁵⁸ TS
<i>Cricetulus griseus</i>	Rodentia	T ⁵⁸ S

Table S2. Number of residues changed and total number of residues. Information of one proBDNF monomer (229 residues), extracted from the program Procheck 3.3.

Residues	number of residues changed	Number of residues
ALA	0	13
CYS	0	6
ASP	0	13
GLU	0	16
PHE	0	4
GLY	0	17
HIS	0	4
ILE	0	8
LYS	0	16
LEU	0	19
MET	0	7
ASN	0	9
PRO	0	10
GLN	0	7
ARG	0	19
SER	0	17
THR	0	17
VAL	0	16
TRP	0	3
TYR	0	8

Table S3. Modeling information on one proBDNF monomer (229 residues), extracted from the program Procheck 3.3.

Parameter	Results
Average value of CA – N –C – CB	34.52 ± 1.32
Distorted main-chain bonds	2
Distorted main-chain angles	1
Distorted planar group	0
Neighbouring contacts	771
Number of hydrogen bonds	126
Potential contact atoms	1803
unusual CA-CA distance for contact number	389
distance	16.1
Phi-psi and chi1-chi2 distributions	
Ramachandran Plot Statistics (for 200 residues)	
[A,B,L] ¹	89.0%
[a,b,l,p] ²	9.5%
[~a,~b,~l,~p] ³	1.5%
[XX] ⁴	0.0%
Number of end-residues (excl. Gly and Pro)	2
Number of glycine residues	17
Number of proline residues	10

¹ Residues in most favored regions

² Residues in additional allowed regions

³ Residues in generously allowed regions

⁴ Residues in disallowed region

Table S4. Sequences of Data Set 1 analyzed in this study.

Order	Species	GenBank accession no.	Length (bp)	Reference
Carnivora	<i>Ailuropoda melanoleuca</i>	U56638	744	Unpublished
Carnivora	<i>Ailurus fulgens</i>	U56639	744	Unpublished
Carnivora	<i>Canis familiaris</i>	XM_850457	1144	Annotation
Carnivora	<i>Felis catus</i>	NM_001009828	744	<i>Murphy et al, 2001a</i>
Carnivora	<i>Helarctos malayanus</i>	AF002240	744	Direct submission
Carnivora	<i>Procyon lotor</i>	AF003188	744	Direct submission
Carnivora	<i>Selenarctos thibetanus</i>	DQ093584	744	Unpublished
Carnivora	<i>Ursus arctos</i>	AF002239	744	Direct submission
Cetartiodactyla	<i>Bos taurus</i>	NM_001046607	1207	<i>Arab et al., 1997</i>
Cetartiodactyla	<i>Lipotes vexillifer</i>	AY700119	747	Direct submission
Cetartiodactyla	<i>Sus scrofa</i>	NM_214259	1184	<i>Leibrok et al., 1989</i>
Didelphimorphia	<i>Monodelphis domestica</i>	XM_001368353	786	Vischer, 1997
Monotremata	<i>Ornithorhynchus anatinus</i>	XM_001511930	819	Genome
Perissodactyla	<i>Equus caballus</i>	NM_001081787	744	<i>Murphy et al, 2001a</i>
Primates	<i>Homo sapiens</i>	EF689015	4016	<i>Pruunsild et al., 2007</i>
Primates	<i>Pan troglodytes</i>	NM_001012441	744	<i>Dorus et al., 2004</i>
Rodentia	<i>Cavia porcellus</i>	AB012097	768	Direct submission
Rodentia	<i>Mus musculus</i>	NM_001048142	4009	<i>Maisonpierre et al., 1990</i>
Rodentia	<i>Rattus norvegicus</i>	BC087634	1625	<i>Strausberg et al., 2002</i>
Rodentia	<i>Spermophilus citellus</i>	AY646114	925	Unpublished

Table S5. Additional sequences included in Data Set 2 (analyzed jointly with those listed on Table S4).

Order	Species	GenBank accession no.	Length (bp)	Reference
Cetartiodactyla	<i>Megaptera novaeangliae</i>	AY011486	561	Murphy et al, 2001a
Cetartiodactyla	<i>Tursiops truncatus</i>	AY011487	561	Murphy et al, 2001a
Cetartiodactyla	<i>Okapia johnstoni</i>	AY011492	567	Murphy et al, 2001a
Cetartiodactyla	<i>Hippopotamus amphibius</i>	AY011488	561	Murphy et al, 2001a
Cetartiodactyla	<i>Tragelaphus eurycerus</i>	AY011490	567	Murphy et al, 2001a
Cetartiodactyla	<i>Lama glama</i>	AY011489	558	Murphy et al, 2001a
Carnivora	<i>Canis simensis</i>	DQ240363	489	Pollinger et al., 2005
Perissodactyla	<i>Tapirus indicus</i>	AY011495	558	Murphy et al, 2001a
Carnivora	<i>Leopardus pardalis</i>	AY011497	558	Murphy et al, 2001a
Carnivora	<i>Panthera onca</i>	AY011498	558	Murphy et al, 2001a
Pholidota	<i>Manis pentadactyla</i>	AY011501	558	Murphy et al, 2001a
Eulipotyphla	<i>Condylura cristata</i>	AY011449	555	Murphy et al, 2001a
Eulipotyphla	<i>Sorex araneus</i>	AY011450	558	Murphy et al, 2001a
Eulipotyphla	<i>Talpa altaica</i>	AY011448	555	Murphy et al, 2001a
Eulipotyphla	<i>Erinaceus concolor</i>	AY011447	555	Murphy et al, 2001a
Eulipotyphla	<i>Solenodon paradoxus</i>	AY530070	558	Roca et al., 2004
Chiroptera	<i>Craseonycteris thonglongyai</i>	AY834531	561	Teeling et al., 2005
Chiroptera	<i>Cynopterus brachyotis</i>	AY834508	558	Teeling et al., 2005
Chiroptera	<i>Emballonura atrata</i>	AY834514	558	Teeling et al., 2005
Chiroptera	<i>Hipposideros commersoni</i>	AY834511	561	Teeling et al., 2005
Chiroptera	<i>Miniopterus fraterculus</i>	EF397711	558	Miller-Butterworth et al., 2007
Chiroptera	<i>Natalus stramineus</i>	AY834529	558	Teeling et al., 2005
Chiroptera	<i>Megaderma lyra</i>	AY059688	561	Murphy et al., 2001b
Chiroptera	<i>Myotis daubentoni</i>	AY834523	558	Teeling et al., 2005
Chiroptera	<i>Eptesicus fuscus</i>	EF397712	558	Miller-Butterworth et al., 2007
Chiroptera	<i>Noctilio albiventris</i>	AY834520	533	Teeling et al., 2005
Chiroptera	<i>Pteronotus parnellii</i>	AY834525	557	Teeling et al., 2005
Chiroptera	<i>Pteropus giganteus</i>	AY011483	558	Murphy et al, 2001a
Chiroptera	<i>Artibeus jamaicensis</i>	AY011482	558	Murphy et al, 2001a
Dermoptera	<i>Cynocephalus variegatus</i>	AY011474	558	Murphy et al, 2001a
Chiroptera	<i>Nycteris thebaica</i>	AY011485	558	Murphy et al, 2001a
Chiroptera	<i>Tadarida brasiliensis</i>	AY059687	558	Murphy et al., 2001b
Chiroptera	<i>Rousettus lanosus</i>	AY011484	558	Murphy et al, 2001a
Chiroptera	<i>Rhogeessa tumida</i>	AY834522	555	Teeling et al., 2005

Table S5. Continued.

Order	Species	GenBank accession no.	Length (bp)	Reference
Chiroptera	<i>Furipterus horrens</i>	AY834528	539	Teeling <i>et al.</i> , 2005
Chiroptera	<i>Myzopoda aurita</i>	AY834524	542	Teeling <i>et al.</i> , 2005
Chiroptera	<i>Rhinolophus creaghi</i>	AY834510	558	Teeling <i>et al.</i> , 2005
Rodentia	<i>Agouti taczanowskii</i>	AY011471	585	Murphy <i>et al.</i> , 2001a
Rodentia	<i>Dinomys branickii</i>	AY011470	558	Murphy <i>et al.</i> , 2001a
Rodentia	<i>Hydrochaeris hydrochaeris</i>	AY011469	561	Murphy <i>et al.</i> , 2001a
Rodentia	<i>Erethizon dorsatum</i>	AY011465	564	Murphy <i>et al.</i> , 2001a
Rodentia	<i>Tamias striatus</i>	AY011457	558	Murphy <i>et al.</i> , 2001a
Rodentia	<i>Muscardinus avellanarius</i>	AY011459	561	Murphy <i>et al.</i> , 2001a
Rodentia	<i>Hystrix brachyurus</i>	AY011464	573	Murphy <i>et al.</i> , 2001a
Rodentia	<i>Castor canadensis</i>	AY011458	558	Murphy <i>et al.</i> , 2001a
Rodentia	<i>Dipodomys heermanni</i>	AY011466	558	Murphy <i>et al.</i> , 2001a
Rodentia	<i>Cricetulus griseus</i>	AY011463	561	Murphy <i>et al.</i> , 2001a
Lagomorpha	<i>Ochotona hyperborea</i>	AY011473	558	Murphy <i>et al.</i> , 2001a
Lagomorpha	<i>Sylvilagus floridanus</i>	AY011472	558	Murphy <i>et al.</i> , 2001a
Primates	<i>Ateles fusciceps</i>	AY011477	561	Murphy <i>et al.</i> , 2001a
Primates	<i>Callimico goeldii</i>	AY011480	561	Murphy <i>et al.</i> , 2001a
Primates	<i>Hylobates concolor</i>	AY011479	558	Murphy <i>et al.</i> , 2001a
Primates	<i>Otolemur garnettii</i>	AF535049	558	Unpublished
Primates	<i>Lemur catta</i>	AY011476	558	Murphy <i>et al.</i> , 2001a
Primates	<i>Macaca mulatta</i>	AY011478	558	Murphy <i>et al.</i> , 2001a
Dermoptera	<i>Cynocephalus volans</i>	AF535048	558	Unpublished
Scandentia	<i>Tupaia minor</i>	AY011475	552	Murphy <i>et al.</i> , 2001a
Scandentia	<i>Ptilocercus lowii</i>	EU142162	555	Janecka <i>et al.</i> , 2007
Scandentia	<i>Urogale everetti</i>	AF535047	558	Unpublished
Xenarthra	<i>Euphractus sexcinctus</i>	AY011443	558	Murphy <i>et al.</i> , 2001a
Xenarthra	<i>Chaetophractus villosus</i>	AY011444	558	Murphy <i>et al.</i> , 2001a
Xenarthra	<i>Choloepus didactylus</i>	AY011442	558	Murphy <i>et al.</i> , 2001a
Xenarthra	<i>Choloepus hoffmanni</i>	AY011441	558	Murphy <i>et al.</i> , 2001a
Xenarthra	<i>Myrmecophaga tridactyla</i>	AY011446	558	Murphy <i>et al.</i> , 2001a
Xenarthra	<i>Tamandua tetradactyla</i>	AY011445	558	Murphy <i>et al.</i> , 2001a
Macroscelidea	<i>Elephantulus rufescens</i>	AY011455	558	Murphy <i>et al.</i> , 2001a

Table S5. Continued.

Order	Species	GenBank accession no.	Length (bp)	Reference
Macroscelidea	<i>Macroscelides proboscideus</i>	AY011454	558	Murphy <i>et al</i> , 2001a
Afrosoricida	<i>Amblysomus hottentotus</i>	AY059685	558	Murphy <i>et al.</i> , 2001b
Sirenia	<i>Trichechus manatus</i>	AY011451	555	Murphy <i>et al</i> , 2001a
Proboscídea	<i>Loxodonta africana</i>	AY011453	558	Murphy <i>et al</i> , 2001a
Hyracoidea	<i>Procavia capensis</i>	AY011452	561	Murphy <i>et al</i> , 2001a
Tubulidentata	<i>Orycteropus afer</i>	AY011456	558	Murphy <i>et al</i> , 2001a
Diprotodontia	<i>Macropus eugenii</i>	AY011440	555	Murphy <i>et al</i> , 2001a
Didelphimorphia	<i>Didelphis virginiana</i>	AY011439	555	Murphy <i>et al</i> , 2001a

Table S6. Sequences of the order *Carnivora* analyzed in this study.

Species	Coverage	GenBank accession no.	Length (bp)	Reference
<i>Ailuropoda melanoleuca</i>	complete cds	U56638	744	Unpublished
<i>Ailurus fulgens</i>	complete cds	U56639	744	Unpublished
<i>Aonyx capensis</i>	partial cds	EF987606	548	Koepfli <i>et al.</i> , 2008
<i>Aonyx cinérea</i>	partial cds	EF987607	548	Koepfli <i>et al.</i> , 2008
<i>Arctonyx collaris</i>	partial cds	EF987639	548	Koepfli <i>et al.</i> , 2008
<i>Bassaricyon alleni</i>	partial cds	DQ660195	548	Koepfli <i>et al.</i> , 2007
<i>Bassaricyon gabbii</i>	partial cds	DQ660196	548	Koepfli <i>et al.</i> , 2007
<i>Bassariscus astutus</i>	partial cds	DQ660197	548	Koepfli <i>et al.</i> , 2007
<i>Bassariscus sumichrasti</i>	partial cds	DQ660198	546	Koepfli <i>et al.</i> , 2007
<i>Canis adustus</i>	partial cds	DQ240358	489	Pollinger <i>et al.</i> , 2005
<i>Canis aureus</i>	partial cds	DQ240359	489	Pollinger <i>et al.</i> , 2005
<i>Canis familiaris</i>	complete cds	XM_850457	1144	Annotation
<i>Canis latrans</i>	partial cds	DQ240360	489	Pollinger <i>et al.</i> , 2005
<i>Canis lúpus</i>	partial cds	DQ240361	489	Pollinger <i>et al.</i> , 2005
<i>Canis simensis</i>	partial cds	DQ240363	489	Pollinger <i>et al.</i> , 2005
<i>Cuon alpinus</i>	partial cds	DQ240366	489	Pollinger <i>et al.</i> , 2005
<i>Dusicyon thous</i>	partial cds	DQ240364	488	Pollinger <i>et al.</i> , 2005
<i>Eira Barbara</i>	partial cds	DQ660206	548	Koepfli <i>et al.</i> , 2007
<i>Enhydra lutris</i>	partial cds	DQ660204	548	Koepfli <i>et al.</i> , 2007
<i>Felis catus</i>	Complete cds	NM_001009828	744	Murphy <i>et al.</i> , 2001a
<i>Galictis cuja</i>	partial cds	EF987638	548	Koepfli <i>et al.</i> , 2008
<i>Galictis vittata</i>	partial cds	EF987637	548	Koepfli <i>et al.</i> , 2008
<i>Gulo gulo</i>	partial cds	EF987635	548	Koepfli <i>et al.</i> , 2008
<i>Helarctos malayanus</i>	complete cds	AF002240	744	Direct submission
<i>Hydrictis maculicollis</i>	partial cds	EF987612	548	Koepfli <i>et al.</i> , 2008
<i>Ictonyx striatus</i>	partial cds	EF987636	548	Koepfli <i>et al.</i> , 2008
<i>Leopardus pardalis</i>	partial cds	AY011497	558	Murphy <i>et al.</i> , 2001a
<i>Lontra canadensis</i>	partial cds	EF987608	548	Koepfli <i>et al.</i> , 2008
<i>Lontra felina</i>	partial cds	EF987609	548	Koepfli <i>et al.</i> , 2008
<i>Lontra longicaudis</i>	partial cds	EF987610	548	Koepfli <i>et al.</i> , 2008
<i>Lutra lutra</i>	partial cds	EF987611	548	Koepfli <i>et al.</i> , 2008
<i>Lutra sumatrana</i>	partial cds	EF987614	548	Koepfli <i>et al.</i> , 2008
<i>Lutrogale perspicillata</i>	partial cds	EF987615	548	Koepfli <i>et al.</i> , 2008
<i>Lycaon pictus</i>	partial cds	DQ240367	489	Pollinger <i>et al.</i> , 2005

Table S6. Continued.

Species	Coverage	GenBank accession no.	Length (bp)	Reference
<i>Martes americana</i>	partial cds	DQ660207	548	Koepfli <i>et al.</i> , 2007
<i>Martes flavigula</i>	partial cds	EF987629	548	Koepfli <i>et al.</i> , 2008
<i>Martes foina</i>	partial cds	EF987630	548	Koepfli <i>et al.</i> , 2008
<i>Martes martes</i>	partial cds	EF987631	548	Koepfli <i>et al.</i> , 2008
<i>Martes melampus</i>	partial cds	EF987632	548	Koepfli <i>et al.</i> , 2008
<i>Martes pennanti</i>	partial cds	EF987633	548	Koepfli <i>et al.</i> , 2008
<i>Martes zibellin</i>	partial cds	EF987634	548	Koepfli <i>et al.</i> , 2008
<i>Meles meles</i>	partial cds	EF987641	548	Koepfli <i>et al.</i> , 2008
<i>Mellivora capensis</i>	partial cds	EF987641	548	Koepfli <i>et al.</i> , 2008
<i>Melogale moschata</i>	partial cds	EF987642	548	Koepfli <i>et al.</i> , 2008
<i>Melogale personata</i>	partial cds	EF987643	548	Koepfli <i>et al.</i> , 2008
<i>Mustela erminea</i>	partial cds	EF987619	548	Koepfli <i>et al.</i> , 2008
<i>Mustela eversmannii</i>	partial cds	EF987620	548	Koepfli <i>et al.</i> , 2008
<i>Mustela frenata</i>	partial cds	EF987621	548	Koepfli <i>et al.</i> , 2008
<i>Mustela nigripes</i>	partial cds	EF987623	548	Koepfli <i>et al.</i> , 2008
<i>Mustela nivalis</i>	partial cds	EF987624	548	Koepfli <i>et al.</i> , 2008
<i>Mustela nudipes</i>	partial cds	EF987625	548	Koepfli <i>et al.</i> , 2008
<i>Mustela putorius</i>	partial cds	EF987626	548	Koepfli <i>et al.</i> , 2008
<i>Mustela sibirica</i>	partial cds	EF987627	548	Koepfli <i>et al.</i> , 2008
<i>Mustela strigidorsa</i>	partial cds	EF987628	548	Koepfli <i>et al.</i> , 2008
<i>Mustela vison</i>	partial cds	DQ660205	548	Koepfli <i>et al.</i> , 2007
<i>Nasua narica</i>	partial cds	DQ660199	548	Koepfli <i>et al.</i> , 2007
<i>Nasua nasua</i>	partial cds	DQ660200	548	Koepfli <i>et al.</i> , 2007
<i>Nyctereutes procyonoides</i>	partial cds	DQ240368	489	Pollinger <i>et al.</i> , 2005
<i>Otocyon megalotis</i>	partial cds	DQ240369	489	Pollinger <i>et al.</i> , 2005
<i>Panthera onça</i>	partial cds	AY011498	558	Murphy <i>et al.</i> , 2001a
<i>Poecilogale albinucha</i>	partial cds	EF987616	548	Koepfli <i>et al.</i> , 2008
<i>Potos flavus</i>	partial cds	DQ660201	548	Koepfli <i>et al.</i> , 2007
<i>Procyon cancrivorus</i>	partial cds	DQ660202	548	Koepfli <i>et al.</i> , 2007
<i>Procyon lotor</i>	complete cds	AF003188	744	Direct submission
<i>Pseudalopex culpaeus</i>	partial cds	DQ240370	489	Pollinger <i>et al.</i> , 2005
<i>Pseudalopex fulvipes</i>	partial cds	DQ240371	479	Pollinger <i>et al.</i> , 2005
<i>Pseudalopex griseus</i>	partial cds	DQ240372	479	Pollinger <i>et al.</i> , 2005

Table S6. Continued.

Species	Coverage	GenBank accession no.	Length (bp)	Reference
<i>Pseudalopex sechurae</i>	partial cds	DQ240374	489	Pollinger <i>et al.</i> , 2005
<i>Pseudalopex vetulus</i>	partial cds	DQ240375	489	Pollinger <i>et al.</i> , 2005
<i>Pteronura brasiliensis</i>	partial cds	EF987613	548	Koepfli <i>et al.</i> , 2008
<i>Selenarctos thibetanus</i>	complete cds	DQ093584	744	Unpublished
<i>Speothos venaticus</i>	partial cds	DQ240376	489	Pollinger <i>et al.</i> , 2005
<i>Taxidea taxus</i>	partial cds	DQ660208	548	Koepfli <i>et al.</i> , 2007
<i>Urocyon cinereoargenteus</i>	partial cds	DQ240377	489	Pollinger <i>et al.</i> , 2005
<i>Urocyon littoralis</i>	partial cds	DQ240378	486	Koepfli <i>et al.</i> , 2007
<i>Ursus americanus</i>	partial cds	DQ240386	489	Koepfli <i>et al.</i> , 2007
<i>Ursus arctos</i>	complete cds	AF002239	744	Direct submission
<i>Vulpes cana</i>	partial cds	DQ240379	476	Pollinger <i>et al.</i> , 2005
<i>Vulpes chama</i>	partial cds	DQ240380	477	Pollinger <i>et al.</i> , 2005
<i>Vulpes corsac</i>	partial cds	DQ240381	488	Pollinger <i>et al.</i> , 2005
<i>Vulpes macrotis</i>	partial cds	DQ240382	489	Pollinger <i>et al.</i> , 2005
<i>Vulpes rueppellii</i>	partial cds	DQ240383	476	Pollinger <i>et al.</i> , 2005
<i>Vulpes vulpes</i>	cds parciais	DQ240384	489	Pollinger <i>et al.</i> , 2005
<i>Vulpes zerda</i>	partial cds	DQ240385	479	Pollinger <i>et al.</i> , 2005

Table S7. Sequences of the order *Cetartiodactyla* analyzed in this study.

Species	Coverage	GenBank accession no.	Length (bp)	Reference
<i>Balaena mysticetus</i>	partial cds	EU444888	569	Demere <i>et al.</i> , 2008
<i>Balaenoptera acutorostrata</i>	partial cds	EU444885	569	Demere <i>et al.</i> , 2008
<i>Balaenoptera bonaerensis</i>	partial cds	EU444884	569	Demere <i>et al.</i> , 2008
<i>Balaenoptera borealis</i>	partial cds	EU444882	569	Demere <i>et al.</i> , 2008
<i>Balaenoptera edeni</i>	partial cds	EU444883	569	Demere <i>et al.</i> , 2008
<i>Balaenoptera musculus</i>	partial cds	EU444881	569	Demere <i>et al.</i> , 2008
<i>Balaenoptera physalus</i>	partial cds	EU444880	569	Demere <i>et al.</i> , 2008
<i>Bos taurus</i>	complete cds	NM_001046607	1207	Arab <i>et al.</i> , 1997
<i>Caperea marginata</i>	partial cds	EU444887	569	Demere <i>et al.</i> , 2008
<i>Delphinus tropicalis</i>	partial cds	AY954641	515	Unpublished
<i>Eschrichtius robustus</i>	partial cds	EU444886	569	Demere <i>et al.</i> , 2008
<i>Eubalaena japonica</i>	partial cds	EU444889	569	Demere <i>et al.</i> , 2008
<i>Hippopotamus amphibius</i>	partial cds	AY011488	561	Murphy <i>et al.</i> , 2001a
<i>Lama glama</i>	partial cds	AY011489	558	Murphy <i>et al.</i> , 2001a
<i>Lipotes vexillifer</i>	complete cds	AY700119	747	Direct submission
<i>Megaptera novaeangliae</i>	partial cds	AY011486	561	Murphy <i>et al.</i> , 2001a
<i>Mesoplodon densirostris</i>	partial cds	AY954636	515	Unpublished
<i>Neophocaena phocaenoides</i>	partial cds	AY954644	515	Unpublished
<i>Okapia johnstoni</i>	partial cds	AY011492	567	Murphy <i>et al.</i> , 2001a
<i>Physeter catodon</i>	partial cds	EU444877	569	Demere <i>et al.</i> , 2008
<i>Pontoporia blainvillei</i>	partial cds	AY954645	515	Unpublished
<i>Sousa chinensis</i>	partial cds	AY954638	515	Unpublished
<i>Stenella coeruleoalba</i>	partial cds	AY954639	515	Unpublished
<i>Sus scrofa</i>	Complete cds	NM_214259	1184	Leibrok <i>et al.</i> , 1989
<i>Tragelaphus eurycerus</i>	partial cds	AY011490	567	Murphy <i>et al.</i> , 2001a
<i>Tursiops aduncus</i>	partial cds	AY954637	515	Unpublished
<i>Tursiops truncatus</i>	partial cds	AY011487	561	Murphy <i>et al.</i> , 2001a
<i>Ziphius cavirostris</i>	partial cds	EU444878	569	Demere <i>et al.</i> , 2008

Table S8. Sequences of the order *Chiroptera* analyzed in this study.

Species	Coverage	GenBank accession no.	Length (bp)	Reference
<i>Anoura geoffroyi</i>	partial cds	AY834519	544	Teeling <i>et al.</i> , 2005
<i>Artibeus jamaicensis</i>	partial cds	AY011482	558	Murphy <i>et al.</i> , 2001a
<i>Craseonycteris thonglongyai</i>	partial cds	AY834531	561	Teeling <i>et al.</i> , 2005
<i>Cynopterus brachyotis</i>	partial cds	AY834508	558	Teeling <i>et al.</i> , 2005
<i>Emballonura atrata</i>	partial cds	AY834514	558	Teeling <i>et al.</i> , 2005
<i>Eptesicus fuscus</i>	partial cds	EF397712	558	Miller-Butterworth <i>et al.</i> , 2007
<i>Eumops auripendulus</i>	partial cds	AY834530	527	Teeling <i>et al.</i> , 2005
<i>Furipterus horrens</i>	partial cds	AY834528	539	Teeling <i>et al.</i> , 2005
<i>Hipposideros commersoni</i>	partial cds	AY834511	561	Teeling <i>et al.</i> , 2005
<i>Macroderma gigas</i>	partial cds	AY834512	542	Teeling <i>et al.</i> , 2005
<i>Megaderma lyra</i>	partial cds	AY059688	561	Murphy <i>et al.</i> , 2001b
<i>Miniopterus fraterculus</i>	partial cds	EF397711	558	Miller-Butterworth <i>et al.</i> , 2007
<i>Miniopterus natalensis</i>	partial cds	EF397710	530	Miller-Butterworth <i>et al.</i> , 2007
<i>Myotis daubentoni</i>	partial cds	AY834523	558	Teeling <i>et al.</i> , 2005
<i>Mystacina tuberculata</i>	partial cds	AY834527	544	Teeling <i>et al.</i> , 2005
<i>Myzopoda aurita</i>	partial cds	AY834524	542	Teeling <i>et al.</i> , 2005
<i>Natalus stramineus</i>	partial cds	AY834529	558	Teeling <i>et al.</i> , 2005
<i>Noctilio albiventris</i>	partial cds	AY834520	533	Teeling <i>et al.</i> , 2005
<i>Pteronotus parnellii</i>	partial cds	AY834525	557	Teeling <i>et al.</i> , 2005
<i>Pteropus giganteus</i>	partial cds	AY011483	558	Murphy <i>et al.</i> , 2001a
<i>Rhinolophus creaghi</i>	partial cds	AY834510	558	Teeling <i>et al.</i> , 2005
<i>Rhinopoma hardwicke</i>	partial cds	AY834513	543	Teeling <i>et al.</i> , 2005
<i>Rhogeessa tumida</i>	partial cds	AY834522	555	Teeling <i>et al.</i> , 2005
<i>Rhynchonycteris naso</i>	partial cds	AY834516	536	Teeling <i>et al.</i> , 2005
<i>Rousettus lanosus</i>	partial cds	AY011484	558	Murphy <i>et al.</i> , 2001a
<i>Tadarida brasiliensis</i>	partial cds	AY059687	558	Murphy <i>et al.</i> , 2001a
<i>Taphozous nudiventris</i>	partial cds	AY834515	538	Teeling <i>et al.</i> , 2005
<i>Thyroptera tricolor</i>	partial cds	AY834526	545	Teeling <i>et al.</i> , 2005

Table S9. Sequences of the order *Rodentia* analyzed in this study.

Species	Coverage	GenBank accession no.	Length (bp)	Reference
<i>Agouti taczanowskii</i>	partial cds	AY011471	585	Murphy <i>et al</i> , 2001a
<i>Castor canadensis</i>	partial cds	AY011458	558	Murphy <i>et al</i> , 2001a
<i>Cavia porcellus</i>	complete cds	AB012097	768	Direct submission
<i>Cavia tschudii</i>	partial cds	AY011468	581	Murphy <i>et al</i> , 2001a
<i>Cricetulus griseus</i>	partial cds	AY011463	561	Murphy <i>et al</i> , 2001a
<i>Dinomys branickii</i>	partial cds	AY011470	558	Murphy <i>et al</i> , 2001a
<i>Dipodomys heermanni</i>	partial cds	AY011466	558	Murphy <i>et al</i> , 2001a
<i>Erethizon dorsatum</i>	partial cds	AY011465	564	Murphy <i>et al</i> , 2001a
<i>Hydrochaeris hydrochaeris</i>	partial cds	AY011469	561	Murphy <i>et al</i> , 2001a
<i>Hystrix brachyurus</i>	partial cds	AY011464	573	Murphy <i>et al</i> , 2001a
<i>Mus musculus</i>	complete cds	NM_001048142	4009	Maisonpierre <i>et al.</i> , 1990
<i>Muscardinus avellanarius</i>	partial cds	AY011459	561	Murphy <i>et al</i> , 2001a
<i>Rattus norvegicus</i>	complete cds	BC087634	1625	Strausberg <i>et al.</i> , 2002
<i>Spermophilus citellus</i>	complete cds	AY646114	925	Unpublished
<i>Tamias striatus</i>	partial cds	AY011457	558	Murphy <i>et al</i> , 2001a

Supplementary references for tables S4-S9.

- Arab, S.; Krohn, K.; Lachmund, A., et al. The gene encoding bovine brain derived neurotrophic factor (BDNF), *Gene*, v.185, pp.95-8, 1997.
- Demere, T.; McGowen, M.; Berta, A.; Gatesy, J. Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. *Syst. Biol.*, v. 57 n.1, pp. 15-37, 2008.
- Dorus, S.; Vallender, E.; Evans, P., et al. Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell*, v.119, n. 7, pp.1027-40, 2004.
- Janecka, J. ; Miller, W.; Pringle, T. et al. Molecular and genomic data identify the closest living relative of primates. *Science*, v. 318, v. 5851, pp. 792-94, 2007.
- Koepfli, K.; Gompper, M.; Eizirik, E., et al. Phylogeny of the Procyonidae (Mammalia: Carnivora): Molecules, morphology and the Great American Interchange. *Mol. Phylogenet. Evol.*, v 43 , n.3, pp.1076-95, 2007.
- Koepfli, K.; Deere, K.; Slater, G. et al. Multigene phylogeny of the Mustelidae: resolving relationships, tempo and biogeographic history of a mammalian adaptive radiation. *BMC Biol.*, v. 6, v.1, pp. 10, 2008.
- Leibrock, J.; Lottspeich, F.; Hohn, A., et al. Molecular cloning and expression of brain derived neurotrophic factor, *Nature.*, v.341, n.6238, pp.149-52, 1989.
- Maisonpierre, P.; Belluscio, L.; Friedman, B., et al. NT-3, BDNF, and NGF in the developing rat nervous system: parallel as well as reciprocal patterns of expression. *Neuron*, v. 5, n. 4, pp. 501-09, 1990.
- Miller-Butterworth, C.; Murphy, W.; O'Brien, S., et al. A Family Matter: Conclusive Resolution of the Taxonomic Position of the Long-fingered Bats, *Miniopterus*. *Mol. Biol. Evol.*, v. 24, n.7, pp.1553-61, 2007.
- Murphy, W.J.; Eizirik, E.; O'Brien, S.J, et al. Resolution of the early placental mammal radiation using Bayesian phylogenetics, *Science*, v.294 , n.5550, pp. 2348-2351, 2001a.
- Murphy, W.; Eizirik, E.; Johnson, W., et al. Molecular phylogenetics and the origins of placental mammals, *Nature*, v.409, n. 6820, pp. 614-8, 2001b.
- Pollinger, J.; Bardeleben, C.; Koepfli, K., et al. Genome Sequence, Comparative Analysis and Haplotype Structure of the Domestic Dog, *Nature*, In press, 2005.
- Pruunsild, P.; Kazantseva, A.; Aid, T., et al. Dissecting the human BDNF locus: bidirectional transcription, complex splicing, and multiple promoters. *Genomics*, v. 90, n.3, pp.397-406, 2007.
- Roca, A.; Bar-Gal, G.; Eizirik, E., et al. Mesozoic origin for West Indian insectivores. *Nature*, v. 429, n. 6992, pp. 649-51, 2004.
- Strausberg, R.; Feingold, E.; Grouse, L. et al. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci USA*, v.99, n. 26, pp.16899-90, 2002.
- Teeling, E.; Springer, M.; Madsen, O., et al. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science*, v. 307, n. 5709, pp. 580-84, 2005.
- Vischer, H. BDNF is expressed at the crush site after spinal cord lesion in newborn opossum (*Monodelphis domestica*). *Eur. J. Neurosci.*, v.9 , n. 9, pp.1993-7,1997.

Table S10. Proportion of conserved sites in different *BDNF* data sets.

	Data set 1	Data set 2	Carnivora	Chiroptera	Cetartiodactyla	Rodentia
Total nucleotide sites	720	570	510	546	543	570
Conserved sites	0.66	0.55	0.77	0.72	0.81	0.71
Total protein sites	240 ^a	190 ^a	170 ^a	182 ^a	181 ^a	183 ^a
	121 ^b	121 ^b	110 ^b	121 ^b	112 ^b	121 ^b
	119 ^c	69 ^c	60 ^c	61 ^c	69 ^c	62 ^c
Conserved sites	0.81 ^a	0.63 ^a	0.83 ^a	0.80 ^a	0.89 ^a	0.81 ^a
	0.68 ^b	0.50 ^b	0.76 ^b	0.71 ^b	0.83 ^b	0.73 ^b
	0.94 ^c	0.83 ^c	0.97 ^c	0.97 ^c	1 ^c	0.98 ^c

^a Considering amino acid sites in the **entire proBDNF**.

^b Considering amino acid sites in the **Pro domain** only.

^c Considering amino acid sites in the **mature domain** only.

Table S11. Parameter Estimates, Log-likelihood Values, and Likelihood Ratio Test for the models M1, M2, M7 and M8. More information about the hypotheses tested can be found on Supplementary Methods. **P**: number of site classes contained in each model, including the ω distribution; **lnL**: log likelihood values; **LRT**: Likelihood Ratio Test (with $\alpha = 0.05$ and $df=1$).

Data set 1.

Model	P	Estimates of Parameters	lnL	Nested Models	LRT	χ^2	Conclusion
M1	2	p1=0.96162; p2=0.03838 w1=0.02472 w2=1	-3107.245				
M2	3	p1=0.96162; p2=0.03838; p3=0 w1=0.02472; w2=1; w3=9.41299	-3107.245	M1-M2	0	3.84	Does not reject H0.
M7	10	p1-p10= 0.10000 w1-w3=0; w4=0.00007; w5=0.0056; w6=0.00284 w7=0.0114; w8=0.03653; w9=0.10849; w10=0.33638	-3100.867	M7-M8	3.752	3.84	Does not reject H0.
M8	11	p1-p10=0.09839; p11=0.01605 w1 e w2=0; w3=0.00006; w4=0.00041; w5=0.00173 w6=0.00545; w7=0.01445; w8=0.03446; w9=0.07947; w10=0.20777; w11=1	-3098.991				

Data set 2.

Model	P	Estimates of Parameters	lnL	Nested Models	LRT	χ^2	Conclusion
M1	2	p1=0.94008; p2=0.05992 w1=0.03814; w2=1	-6780.185				
M2	3	p1=0.94008; p2=0.05992; p3=0 w1=0.03814; w2=1; w3=2.50723	-6780.185	M1-M2	0	3.84	Does not reject H0.
M7	10	p1-p10=0.10000 w1 e w2=0; w3=0.00007; w4=0.00066; w5=0.00360 w6=0.01402; w7=0.04362; w8=0.11619; w9=0.27822; w10=0.63176	-6725.628	M7-M8	22.642	3.84	Accepts H1
M8	11	p1-p10=0.09708; p11=0.02924 w1=0; w2=0.00005; w3=0.00048; w4=0.00203; w5=0.00601; w6=0.01449; w7=0.03082; w8=0.06119; w9=0.12011; w10=0.26712; w11=1	-6714.307				

Table S11. Continued.

Cetartiodactyla data set.

Model	P	Estimates of Parameters	lnL	Nested Models	LRT	χ^2	Conclusion
M1	2	p1=0.96184; p2=0.03816 w1=0.01746; w2=1	-1278.477				
M2	3	p1=0.97142; p2=0.01570; p3=0.01288 w1=0.02134; w2=1; w3=2.08966	-1278.057	M1-M2	0.84	3.84	Does not reject H0.
M7	10	p1-p10=0.10000 w1-w6=0; w7=0.00001; w8=0.00075 w9=0.03431; w10=0.64583	-1279.879	M7-M8	4.32	3.84	Accepts H1
M8	11	p1-p10=0.09841; p11=0.01586 w1 e w2=0; w3=0.00001; w4=0.00007; w5=0.00048; w6=0.00214; w7=0.0755; w8=0.02279; w9=0.06397; w10=0.19908; w11=2.19016	-1277.719				

Chiroptera data set.

Model	P	Estimates of Parameters	lnL	Nested Models	LRT	χ^2	Conclusion
M1	2	p1=0.93457; p2=0.06543 w1=0.02059; w2=1	-2355.118				
M2	3	p1=0.93457; p2=0.06543; p3=0 w1=0.02059; w2=1; w3=35.35224	-2355.118	M1-M2	0	3.84	Does not reject H0.
M7	10	p1-p10=0.10000 w1-w4=0; w5=0.00006; w6=0.00068; w7=0.00512; w8=0.02880; w9=0.131096; w11=0.51289	-2353.245	M7-M8	0.366	3.84	Does not reject H0.
M8	11	p1-p10=0.09731; p11=0.02689 w1- w3=0; w4=0.00004; w5=0.00035; w6=0.00196; w7=0.00816; w8=0.02821; w9=0.08795; w10=0.28925; w11=1	-2353.062				

Table S11. Continued.

Carnivora data set.

Model	P	Estimates of Parameters	lnL	Nested Models	LRT	χ^2	Conclusion
M1	2	p1= 0.96447; p2=0.03553 w1=0.03928; w2=1	-1523.436636				
M2	3	p1= 0.96447; p2= 0.02060; p3= 0.01493 w1= 0.03928; w2=1; w3=1	-1523.436636	M1-M2	0	3.84	Does not reject H0.
M7	10	p1-p10=0.10000 w1=0;w2= 0.00001; w3= 0.00013 w4= 0.00087; w5= 0.00355; w6= 0.01101; w7= 0.02866; w8= 0.06699; w9= 0.14964; w10= 0.36281	-1521.929454	M7-M8	0.000812	3.84	Does not reject H0.
M8	11	p1-p10=10000; p11= 0.00001 w1=0; w2= 0.00001; w3= 0.00013; w4= 0.00087; w5= 0.00355; w6= 0.01101; w7= 0.02867; w8= 0.06699; w9= 0.14963; w10= 0.36277; w11= 1	-1521.929860				

Rodentia data set.

Model	P	Estimates of Parameters	lnL	Nested Models	LRT	χ^2	Conclusion
M1	2	p1=0.94095; p2=0.05905 w1=0.03053; w2=1	-2182.025				
M2	3	p1=0.94095; p2=0.05905; p3=0 w1=0.03053; w2=1; w3=70.11957	-2182.025	M1-M2	0	3.84	Does not reject H0.
M7	10	p1-p10=0.10000 w1-w3=0; w4=0.00007; w5=0.00058; w6=0.00335; w7=0.01435; w8=0.05047; w9=0.15613; w10=0.46989	-2175.722	M7-M8	2.798	3.84	Does not reject H0.
M8	11	p1-p10=0.09888; p11=0.01118 w1 e w2=0; w3=0.00002; w4=0.00024; w5=0.00134; w6=0.00536; w7=0.01721; w8=0.04785; w9=0.12381; w10=0.34213; w11=1.62736	-2174.323				

Table S12. Analysis of natural selection in the BDNF coding sequence (sites 1-18 correspond to the signal peptide and were excluded), based on data sets 1 and 2 (see Supplementary Methods). For each residue, we indicate the posterior probability of belonging to the site class exhibiting evidence of positive selection, as well as the mean \pm standard error (SE) of the nonsynonymous to synonymous substitution ratio (ω). Sites indicating very low probability of positive selection (i.e. close to zero) indicate high probability of negative selection (i.e. functional constraint). Residues are numbered according to the human sequence. Sites inferred to be under positive selection (see Table 1) are marked with an asterisk. Sites experimentally demonstrated to be involved in specific aspects of BDNF function are highlighted in bold, and detailed in the footnotes.

Residue	Data set 1		Data set 2	
	Probab.	$\omega \pm$ SE	Probab.	$\omega \pm$ SE
19A	0.00000	0.055 \pm 0.029	0.00000	0.051 \pm 0.009
20P	0.00002	0.058 \pm 0.040	0.00000	0.051 \pm 0.013
21M	0.00000	0.053 \pm 0.020	0.00000	0.051 \pm 0.011
22K	0.00000	0.056 \pm 0.034	0.00000	0.051 \pm 0.013
23E	0.00001	0.056 \pm 0.035	0.00000	0.054 \pm 0.022
24A	0.00489	0.173 \pm 0.189	0.00000	0.185 \pm 0.125
25N	0.00022	0.121 \pm 0.124	0.00000	0.286 \pm 0.103
26I	0.00884	0.279 \pm 0.220	0.36505	1.021 \pm 0.377
27R	0.03965	0.199 \pm 0.319	0.05587	0.463 \pm 0.326
28G	0.00112	0.090 \pm 0.111	0.00000	0.058 \pm 0.036
29Q	0.05218	0.320 \pm 0.349	0.00006	0.182 \pm 0.146
30G	0.00000	0.054 \pm 0.027	0.00000	0.051 \pm 0.008
31G	0.00008	0.107 \pm 0.108	0.00000	0.276 \pm 0.108
32L	0.00009	0.060 \pm 0.051	0.00004	0.191 \pm 0.147
33A	0.00000	0.055 \pm 0.029	0.00000	0.058 \pm 0.032
34Y	0.00000	0.056 \pm 0.034	0.00000	0.054 \pm 0.022
35P	0.00475	0.113 \pm 0.159	0.00000	0.053 \pm 0.023
36G	0.42198	0.952 \pm 0.530	0.00012	0.266 \pm 0.156
37V	0.00953	0.288 \pm 0.224	0.00000	0.108 \pm 0.082
38R	0.00002	0.058 \pm 0.040	0.00000	0.052 \pm 0.014
39T	0.00000	0.055 \pm 0.029	0.00000	0.051 \pm 0.010
40H	0.00001	0.057 \pm 0.037	0.00000	0.056 \pm 0.028
41G	0.00000	0.054 \pm 0.027	0.00000	0.056 \pm 0.026
42T	0.00165	0.097 \pm 0.123	0.00061	0.303 \pm 0.169
43L	0.00004	0.059 \pm 0.045	0.00000	0.052 \pm 0.015
44E	0.00000	0.053 \pm 0.022	0.00000	0.051 \pm 0.010
45S	0.00000	0.065 \pm 0.053	0.00000	0.051 \pm 0.008
46V	0.07099	0.503 \pm 0.335	0.00159	0.563 \pm 0.145
47N	0.02072	0.332 \pm 0.261	0.00006	0.371 \pm 0.138
48G	0.00000	0.054 \pm 0.026	0.00000	0.051 \pm 0.007
49P	0.00001	0.057 \pm 0.038	0.00000	0.055 \pm 0.026
50K	0.00009	0.104 \pm 0.106	0.00000	0.281 \pm 0.102
51A	0.12969	0.535 \pm 0.429	0.00027	0.336 \pm 0.154
52G	0.00002	0.057 \pm 0.039	0.00000	0.074 \pm 0.063
53S *	0.00765	0.124 \pm 0.182	0.77941	1.329 \pm 0.331
54R ¹	0.00003	0.057 \pm 0.039	0.00000	0.058 \pm 0.035
55G ¹	0.00007	0.074 \pm 0.073	0.00000	0.073 \pm 0.054
56L ¹	0.00004	0.059 \pm 0.044	0.00000	0.072 \pm 0.059
57T ¹	0.00001	0.057 \pm 0.038	0.00000	0.104 \pm 0.089

Table S12. Continued.

Residue	Data set 1		Data set 2	
	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
58S ^{1,2}	0.00147	0.096 \pm 0.122	0.00047	0.222 \pm 0.169
59L ¹	0.00008	0.060 \pm 0.049	0.00000	0.059 \pm 0.038
60A	0.00005	0.059 \pm 0.046	0.00000	0.059 \pm 0.039
61D	0.00000	0.062 \pm 0.046	0.00000	0.051 \pm 0.009
62T	0.00006	0.059 \pm 0.047	0.00000	0.053 \pm 0.021
63F	0.00258	0.104 \pm 0.138	0.00000	0.059 \pm 0.038
64E	0.00000	0.056 \pm 0.032	0.00000	0.051 \pm 0.011
65H	0.00000	0.055 \pm 0.031	0.00000	0.051 \pm 0.009
66V ³	0.00000	0.054 \pm 0.027	0.00000	0.051 \pm 0.008
67I	0.00000	0.055 \pm 0.032	0.00000	0.051 \pm 0.011
68E	0.00000	0.055 \pm 0.032	0.00000	0.051 \pm 0.010
69E	0.00000	0.062 \pm 0.046	0.00000	0.051 \pm 0.010
70L	0.00004	0.059 \pm 0.045	0.00000	0.052 \pm 0.016
71L	0.00007	0.060 \pm 0.047	0.00000	0.052 \pm 0.018
72D	0.00003	0.071 \pm 0.067	0.00000	0.195 \pm 0.110
73E	0.00000	0.053 \pm 0.022	0.00000	0.051 \pm 0.010
74D	0.00004	0.096 \pm 0.096	0.00000	0.053 \pm 0.019
75Q	0.00000	0.056 \pm 0.034	0.00000	0.085 \pm 0.070
76K	0.00133	0.143 \pm 0.151	0.00000	0.065 \pm 0.047
77V	0.01513	0.221 \pm 0.245	0.01350	0.550 \pm 0.193
78R	0.00070	0.085 \pm 0.100	0.00001	0.274 \pm 0.135
79P	0.00001	0.057 \pm 0.038	0.00000	0.068 \pm 0.052
80N [*]	0.52648	1.092 \pm 0.489	0.59787	1.205 \pm 0.369
81E	0.00000	0.055 \pm 0.029	0.00000	0.058 \pm 0.032
82E	0.00000	0.054 \pm 0.027	0.00000	0.053 \pm 0.020
83N	0.00000	0.063 \pm 0.048	0.00000	0.283 \pm 0.102
84N [*]	0.72621	1.320 \pm 0.501	0.98596	1.496 \pm 0.101
85K	0.00000	0.053 \pm 0.023	0.00000	0.050 \pm 0.006
86D	0.00000	0.054 \pm 0.023	0.00000	0.050 \pm 0.006
87A	0.00000	0.055 \pm 0.029	0.00000	0.070 \pm 0.051
88D	0.00000	0.053 \pm 0.021	0.00000	0.050 \pm 0.005
89L	0.00100	0.093 \pm 0.113	0.00010	0.266 \pm 0.155
90Y	0.00001	0.057 \pm 0.037	0.00000	0.054 \pm 0.023
91T	0.00001	0.057 \pm 0.036	0.00000	0.142 \pm 0.121
92S	0.00003	0.059 \pm 0.044	0.00000	0.052 \pm 0.015
93R	0.00001	0.056 \pm 0.035	0.00000	0.051 \pm 0.009
94V	0.00000	0.054 \pm 0.027	0.00000	0.052 \pm 0.015
95M	0.00000	0.053 \pm 0.020	0.00000	0.051 \pm 0.011
96L	0.00002	0.058 \pm 0.040	0.00000	0.051 \pm 0.014
97S	0.00000	0.054 \pm 0.026	0.00000	0.051 \pm 0.013
98S	0.00001	0.056 \pm 0.035	0.00000	0.056 \pm 0.031
99Q	0.00012	0.061 \pm 0.053	0.00000	0.053 \pm 0.020
100V	0.00000	0.054 \pm 0.027	0.00000	0.051 \pm 0.008
101P	0.00039	0.064 \pm 0.065	0.00000	0.055 \pm 0.027
102L	0.00007	0.060 \pm 0.049	0.00000	0.053 \pm 0.019
103E	0.00000	0.053 \pm 0.022	0.00000	0.050 \pm 0.006
104P	0.00011	0.060 \pm 0.051	0.00000	0.053 \pm 0.019
105P	0.00012	0.061 \pm 0.052	0.00000	0.053 \pm 0.022
106L	0.00011	0.061 \pm 0.052	0.00000	0.062 \pm 0.047

Table S12. Continued.

Residue	Data set 1		Data set 2	
	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
107L	0.00001	0.057 \pm 0.038	0.00000	0.055 \pm 0.026
108F	0.00009	0.060 \pm 0.051	0.00000	0.053 \pm 0.021
109L	0.00003	0.059 \pm 0.044	0.00000	0.052 \pm 0.015
110L	0.00020	0.062 \pm 0.057	0.00000	0.053 \pm 0.019
111E	0.00000	0.053 \pm 0.022	0.00000	0.050 \pm 0.005
112E	0.00000	0.056 \pm 0.032	0.00000	0.051 \pm 0.011
113Y	0.00000	0.056 \pm 0.034	0.00000	0.051 \pm 0.012
114K	0.00001	0.056 \pm 0.035	0.00000	0.051 \pm 0.013
115N	0.00000	0.056 \pm 0.033	0.00000	0.051 \pm 0.012
116Y	0.00000	0.056 \pm 0.034	0.00000	0.051 \pm 0.012
117L	0.00005	0.059 \pm 0.045	0.00000	0.052 \pm 0.015
118D	0.00000	0.055 \pm 0.032	0.00000	0.051 \pm 0.009
119A	0.00002	0.057 \pm 0.038	0.00000	0.052 \pm 0.015
120A	0.00004	0.059 \pm 0.044	0.00000	0.052 \pm 0.017
121N	0.00000	0.053 \pm 0.021	0.00000	0.050 \pm 0.006
122M	0.00000	0.053 \pm 0.020	0.00000	0.050 \pm 0.006
123S	0.00005	0.059 \pm 0.046	0.00000	0.053 \pm 0.020
124M	0.00000	0.061 \pm 0.043	0.00000	0.050 \pm 0.006
125R ⁴	0.00000	0.054 \pm 0.026	0.00000	0.051 \pm 0.009
126V ⁴	0.00000	0.055 \pm 0.030	0.00000	0.051 \pm 0.009
127R ⁴	0.00003	0.059 \pm 0.043	0.00000	0.052 \pm 0.015
128R ⁴	0.00001	0.057 \pm 0.036	0.00000	0.051 \pm 0.013
129H	0.00000	0.055 \pm 0.029	0.00000	0.051 \pm 0.009
130S	0.00016	0.062 \pm 0.055	0.00000	0.053 \pm 0.022
131D	0.00000	0.053 \pm 0.021	0.00000	0.050 \pm 0.005
132P	0.00004	0.059 \pm 0.044	0.00000	0.052 \pm 0.018
133A	0.00004	0.071 \pm 0.068	0.00000	0.051 \pm 0.009
134R	0.00001	0.057 \pm 0.037	0.00000	0.052 \pm 0.014
135R	0.00003	0.058 \pm 0.042	0.00000	0.052 \pm 0.018
136G	0.00000	0.054 \pm 0.026	0.00000	0.051 \pm 0.008
137E	0.00000	0.053 \pm 0.022	0.00000	0.050 \pm 0.005
138L	0.00004	0.059 \pm 0.044	0.00000	0.052 \pm 0.015
139S	0.00000	0.054 \pm 0.023	0.00000	0.051 \pm 0.009
140V	0.00000	0.054 \pm 0.027	0.00000	0.051 \pm 0.008
141C	0.00005	0.059 \pm 0.046	0.00000	0.053 \pm 0.019
142D	0.00000	0.053 \pm 0.021	0.00000	0.050 \pm 0.005
143S	0.00002	0.069 \pm 0.062	0.00000	0.051 \pm 0.011
144I ⁵	0.00024	0.080 \pm 0.088	0.00000	0.108 \pm 0.093
145S	0.00000	0.054 \pm 0.023	0.00000	0.051 \pm 0.009
146E ⁵	0.00000	0.053 \pm 0.022	0.00000	0.051 \pm 0.010
147W	0.00002	0.058 \pm 0.042	0.00000	0.058 \pm 0.036
148V	0.00000	0.055 \pm 0.030	0.00000	0.051 \pm 0.011
149T	0.00003	0.058 \pm 0.042	0.00000	0.052 \pm 0.015
150A	0.00000	0.055 \pm 0.029	0.00000	0.051 \pm 0.010
151A	0.00000	0.056 \pm 0.032	0.00000	0.051 \pm 0.013
152D	0.00000	0.055 \pm 0.031	0.00000	0.051 \pm 0.010
153K	0.00000	0.056 \pm 0.033	0.00000	0.051 \pm 0.013

Table S12. Continued.

Residues	Data set 1		Data set 2	
	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
154K ⁶	0.00000	0.053 \pm 0.022	0.00000	0.050 \pm 0.006
155T ⁶	0.00004	0.058 \pm 0.043	0.00000	0.053 \pm 0.020
156A ⁶	0.00006	0.059 \pm 0.046	0.00000	0.052 \pm 0.017
157V ⁶	0.00000	0.054 \pm 0.027	0.00000	0.051 \pm 0.008
158D ⁶	0.00000	0.053 \pm 0.021	0.00000	0.050 \pm 0.005
159M ⁶	0.00000	0.053 \pm 0.020	0.00000	0.050 \pm 0.006
160S ⁶	0.00003	0.059 \pm 0.043	0.00000	0.052 \pm 0.018
161G ⁶	0.00000	0.055 \pm 0.029	0.00000	0.051 \pm 0.009
162G ⁶	0.00000	0.054 \pm 0.027	0.00000	0.051 \pm 0.007
163T ⁶	0.00000	0.055 \pm 0.029	0.00000	0.051 \pm 0.010
164V	0.00000	0.055 \pm 0.029	0.00000	0.051 \pm 0.009
165T	0.00001	0.056 \pm 0.034	0.00000	0.051 \pm 0.013
166V	0.00000	0.055 \pm 0.030	0.00000	0.051 \pm 0.009
167L	0.00009	0.060 \pm 0.049	0.00000	0.053 \pm 0.020
168E	0.00000	0.056 \pm 0.032	0.00000	0.051 \pm 0.010
169K	0.00000	0.055 \pm 0.031	0.00000	0.051 \pm 0.010
170V	0.00000	0.055 \pm 0.030	0.00000	0.051 \pm 0.009
171P	0.00006	0.059 \pm 0.047	0.00000	0.053 \pm 0.020
172V	0.00001	0.056 \pm 0.034	0.00000	0.051 \pm 0.013
173S ⁶	0.00116	0.094 \pm 0.116	0.00000	0.057 \pm 0.031
174K ⁶	0.00001	0.056 \pm 0.035	0.00000	0.051 \pm 0.013
175G ⁶	0.00000	0.054 \pm 0.027	0.00000	0.051 \pm 0.008
176K ⁶	0.00009	0.060 \pm 0.050	0.00000	0.052 \pm 0.016
177L ⁶	0.00004	0.059 \pm 0.044	0.00000	0.052 \pm 0.015
178K	0.00000	0.053 \pm 0.022	0.00000	0.050 \pm 0.006
179Q	0.00003	0.058 \pm 0.043	0.00000	0.052 \pm 0.014
180Y	0.00001	0.057 \pm 0.036	0.00000	0.051 \pm 0.012
181F	0.00000	0.056 \pm 0.035	0.00000	0.054 \pm 0.024
182Y	0.00001	0.056 \pm 0.035	0.00000	0.054 \pm 0.022
183E	0.00000	0.053 \pm 0.022	0.00000	0.050 \pm 0.005
184T	0.00000	0.055 \pm 0.030	0.00000	0.051 \pm 0.010
185k	0.00000	0.053 \pm 0.022	0.00000	0.051 \pm 0.011
186C	0.00001	0.057 \pm 0.037	0.00000	0.052 \pm 0.015
187N	0.00000	0.055 \pm 0.031	0.00000	0.051 \pm 0.012
188P	0.00001	0.058 \pm 0.040	0.00000	0.051 \pm 0.014
189M	0.00000	0.053 \pm 0.020	0.00000	0.053 \pm 0.019
190G	0.00002	0.058 \pm 0.041	0.00000	0.056 \pm 0.029
191Y	0.00001	0.057 \pm 0.038	0.00000	0.054 \pm 0.023
192T	0.00003	0.058 \pm 0.042	0.00000	0.052 \pm 0.015
193K	0.00000	0.054 \pm 0.025	0.00000	0.051 \pm 0.008
194E	0.00000	0.054 \pm 0.024	0.00000	0.050 \pm 0.007
195G	0.00000	0.055 \pm 0.030	0.00000	0.051 \pm 0.009
196C	0.00001	0.057 \pm 0.036	0.00000	0.052 \pm 0.016
197R	0.00000	0.054 \pm 0.027	0.00000	0.051 \pm 0.011
198G	0.00000	0.054 \pm 0.027	-	-
199I	0.00000	0.056 \pm 0.033	-	-
200D	0.00000	0.053 \pm 0.021	-	-
201K	0.00000	0.055 \pm 0.029	-	-
202R	0.00000	0.054 \pm 0.024	-	-

Table S12. Continued.

Residues	Data set 1		Data set 2	
	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
203H	0.00001	0.057 \pm 0.038	-	-
204W	0.00002	0.058 \pm 0.042	-	-
205N	0.00000	0.063 \pm 0.048	-	-
206S	0.00001	0.058 \pm 0.040	-	-
207K ⁶	0.00001	0.056 \pm 0.035	-	-
208C ⁶	0.00001	0.057 \pm 0.036	-	-
209R ⁶	0.00157	0.070 \pm 0.091	-	-
210T ⁶	0.00009	0.060 \pm 0.050	-	-
211T ⁶	0.00000	0.055 \pm 0.029	-	-
212Q ⁶	0.00000	0.056 \pm 0.035	-	-
213S ⁶	0.00003	0.059 \pm 0.044	-	-
214Y ⁶	0.00001	0.057 \pm 0.037	-	-
215V ⁶	0.00000	0.055 \pm 0.029	-	-
216R ⁶	0.00002	0.058 \pm 0.041	-	-
217A	0.00000	0.055 \pm 0.029	-	-
218L	0.00004	0.059 \pm 0.044	-	-
219T	0.00000	0.055 \pm 0.030	-	-
220M	0.00000	0.053 \pm 0.020	-	-
221D	0.00000	0.056 \pm 0.033	-	-
222S	0.00001	0.066 \pm 0.057	-	-
223K ⁷	0.00001	0.056 \pm 0.035	-	-
224K ⁷	0.00000	0.053 \pm 0.023	-	-
225R ⁷	0.00018	0.060 \pm 0.053	-	-
226I	0.00059	0.087 \pm 0.101	-	-
227G	0.00000	0.055 \pm 0.028	-	-
228W	0.00002	0.058 \pm 0.042	-	-
229R	0.00007	0.059 \pm 0.045	-	-
230F	0.00000	0.056 \pm 0.035	-	-
231I	0.00000	0.056 \pm 0.033	-	-
232R	0.00000	0.054 \pm 0.027	-	-
233I ⁵	0.00000	0.056 \pm 0.033	-	-
234D ⁵	0.00000	0.053 \pm 0.021	-	-
235T	0.00009	0.060 \pm 0.050	-	-
236S	0.00002	0.058 \pm 0.040	-	-
237C	0.00018	0.062 \pm 0.057	-	-
238V	0.00008	0.060 \pm 0.049	-	-
239C	0.00018	0.062 \pm 0.057	-	-
240T	0.00009	0.060 \pm 0.050	-	-
241L	0.00008	0.060 \pm 0.048	-	-
242T	0.00000	0.055 \pm 0.030	-	-
243I	0.00002	0.058 \pm 0.041	-	-
244K	0.00000	0.055 \pm 0.029	-	-
245R	0.00000	0.054 \pm 0.024	-	-
246G	0.00006	0.059 \pm 0.046	-	-
247R	0.00002	0.058 \pm 0.041	-	-

1. Short proBDNF cleavage motif.
2. Insertion point for the serine/threonine expansions observed in cetartiodactyls and caviomorph rodents (see Table S1 and Figures 1 and S4).
3. Functional variant Val66Met identified in humans.
4. Mature BDNF cleavage motif.
5. Motifs mediating interaction with the carboxipeptidase-E co-receptor.
6. Motifs mediating interaction with the TrkB receptor.
7. Motif mediating interaction with the p75 and TrkB receptors.

Table S13. Analysis of natural selection in the BDNF coding sequence (sites 1-18 correspond to the signal peptide and were excluded), based on taxon-specific data sets (see Supplementary Methods). Statistics and annotations are as described in Table S12. Sites inferred to be under positive selection in at least one of data sets (see Table 1) are marked with an asterisk.

Residues	Carnivora		Cetartiodactyla		Chiroptera		Rodentia	
	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
19A		-		-	0.00001	0.057 \pm 0.037	0.00001	0.057 \pm 0.036
20P		-		-	0.00005	0.060 \pm 0.047	0.00003	0.059 \pm 0.045
21M		-		-	0.00000	0.055 \pm 0.028	0.00000	0.056 \pm 0.032
22K		-		-	0.00008	0.061 \pm 0.051	0.00031	0.064 \pm 0.065
23E		-		-	0.00003	0.059 \pm 0.046	0.00014	0.062 \pm 0.056
24A		-		-	0.00007	0.060 \pm 0.049	0.00128	0.096 \pm 0.120
25N		-		-	0.00003	0.074 \pm 0.073	0.00151	0.161 \pm 0.166
26I		-		-	0.00331	0.178 \pm 0.185	0.20854	0.709 \pm 0.468
27R		-		-	0.00249	0.074 \pm 0.105	0.13811	0.392 \pm 0.557
28G		-	0.16650	0.520 \pm 0.807	0.00053	0.065 \pm 0.071	0.00010	0.061 \pm 0.052
29Q		-	0.01224	0.104 \pm 0.241	0.11094	0.432 \pm 0.439	0.03865	0.205 \pm 0.326
30G	0.00073	0.068 \pm 0.079	0.00072	0.068 \pm 0.082	0.00000	0.057 \pm 0.036	0.00000	0.057 \pm 0.036
31G	0.01630	0.161 \pm 0.241	0.00068	0.068 \pm 0.081	0.00000	0.057 \pm 0.036	0.02488	0.368 \pm 0.274
32L	0.00485	0.082 \pm 0.138	0.00684	0.089 \pm 0.178	0.00031	0.064 \pm 0.063	0.00033	0.064 \pm 0.065
33A	0.19521	0.590 \pm 0.537	0.00100	0.070 \pm 0.089	0.00001	0.057 \pm 0.037	0.00001	0.057 \pm 0.036
34Y	0.00177	0.073 \pm 0.099	0.00197	0.075 \pm 0.108	0.00003	0.059 \pm 0.044	0.00003	0.059 \pm 0.045
35P	0.20126	0.533 \pm 0.764	0.01593	0.113 \pm 0.281	0.00134	0.069 \pm 0.088	0.00067	0.066 \pm 0.075
36G	0.00375	0.079 \pm 0.126	0.00827	0.093 \pm 0.196	0.00042	0.065 \pm 0.067	0.02305	0.171 \pm 0.269
37V	0.00046	0.066 \pm 0.072	0.01660	0.165 \pm 0.260	0.00008	0.060 \pm 0.050	0.00001	0.057 \pm 0.036
38R	0.00209	0.074 \pm 0.104	0.00259	0.077 \pm 0.119	0.00013	0.061 \pm 0.055	0.00009	0.061 \pm 0.053
39T	0.00111	0.070 \pm 0.088	0.00116	0.071 \pm 0.092	0.00001	0.058 \pm 0.040	0.00001	0.057 \pm 0.039
40H	0.00354	0.079 \pm 0.123	0.00698	0.090 \pm 0.181	0.00458	0.117 \pm 0.162	0.00006	0.060 \pm 0.049
41G	0.00038	0.065 \pm 0.069	0.00056	0.067 \pm 0.077	0.00000	0.056 \pm 0.033	0.00019	0.082 \pm 0.089
42T	0.00481	0.082 \pm 0.137	0.01024	0.099 \pm 0.219	0.01550	0.150 \pm 0.230	0.02411	0.173 \pm 0.273
43L	0.00302	0.077 \pm 0.117	0.00451	0.083 \pm 0.148	0.00180	0.102 \pm 0.131	0.00012	0.062 \pm 0.055
44E	0.00019	0.063 \pm 0.061	0.00043	0.066 \pm 0.072	0.00001	0.068 \pm 0.061	0.00000	0.055 \pm 0.031

Table S13. Continued.

Residues	Carnivora		Cetartiodactyla		Chiroptera		Rodentia	
	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
45S	0.00076	0.068 \pm 0.080	0.00068	0.068 \pm 0.081	0.00001	0.057 \pm 0.037	0.00001	0.057 \pm 0.037
46V	0.10198	0.438 \pm 0.434	0.03558	0.216 \pm 0.353	0.07736	0.530 \pm 0.339	0.00337	0.177 \pm 0.185
47N *	0.24152	0.662 \pm 0.578	0.00427	0.082 \pm 0.144	0.61496	1.169 \pm 0.464	0.02816	0.274 \pm 0.296
48G	0.00038	0.065 \pm 0.069	0.00056	0.067 \pm 0.077	0.00000	0.056 \pm 0.033	0.00000	0.057 \pm 0.036
49P	0.00250	0.075 \pm 0.110	0.00287	0.078 \pm 0.123	0.00005	0.060 \pm 0.047	0.00003	0.059 \pm 0.045
50K	0.00023	0.064 \pm 0.063	0.00044	0.066 \pm 0.073	0.00002	0.071 \pm 0.067	0.02699	0.368 \pm 0.280
51A *	0.00613	0.086 \pm 0.150	0.95891	2.423 \pm 1.195	0.07566	0.370 \pm 0.391	0.00476	0.117 \pm 0.164
52G	0.02089	0.172 \pm 0.262	0.00827	0.093 \pm 0.196	0.00033	0.064 \pm 0.064	0.00187	0.102 \pm 0.132
53S *	0.00289	0.077 \pm 0.115	0.01884	0.121 \pm 0.311	0.37821	0.871 \pm 0.535	0.84950	1.636 \pm 0.736
54R¹	0.00719	0.088 \pm 0.160	0.00773	0.092 \pm 0.189	0.00043	0.065 \pm 0.068	0.00043	0.065 \pm 0.068
55G¹	0.01670	0.162 \pm 0.243	0.00072	0.068 \pm 0.082	0.00000	0.057 \pm 0.036	0.00026	0.084 \pm 0.094
56L¹	0.00302	0.077 \pm 0.117	0.00451	0.083 \pm 0.148	0.00007	0.060 \pm 0.050	0.00281	0.109 \pm 0.146
57T¹	0.00410	0.080 \pm 0.130	0.00110	0.071 \pm 0.091	0.00004	0.059 \pm 0.045	0.01467	0.233 \pm 0.250
58S^{1,2}	0.00301	0.077 \pm 0.117	0.00381	0.081 \pm 0.138	0.02350	0.169 \pm 0.264	0.23282	0.637 \pm 0.589
59L¹	0.00485	0.082 \pm 0.138	0.00685	0.089 \pm 0.180	0.00028	0.063 \pm 0.061	0.00014	0.062 \pm 0.056
60A	0.00351	0.079 \pm 0.123	0.00567	0.086 \pm 0.165	0.00037	0.064 \pm 0.065	0.00070	0.067 \pm 0.076
61D	0.00027	0.064 \pm 0.065	0.00043	0.066 \pm 0.072	0.00000	0.054 \pm 0.027	0.00000	0.055 \pm 0.029
62T	0.00481	0.082 \pm 0.137	0.00871	0.095 \pm 0.201	0.00081	0.067 \pm 0.078	0.00063	0.066 \pm 0.074
63F	0.00597	0.085 \pm 0.149	0.18709	0.572 \pm 0.871	0.00054	0.065 \pm 0.071	0.00013	0.062 \pm 0.055
64E	0.00312	0.077 \pm 0.118	0.00374	0.080 \pm 0.137	0.00002	0.059 \pm 0.043	0.00003	0.059 \pm 0.045
65H	0.00104	0.070 \pm 0.086	0.00198	0.074 \pm 0.109	0.00001	0.057 \pm 0.038	0.00001	0.057 \pm 0.038
66V³	0.00047	0.066 \pm 0.072	0.00081	0.069 \pm 0.084	0.00000	0.056 \pm 0.033	0.00001	0.057 \pm 0.036
67I	0.00245	0.075 \pm 0.109	0.00084	0.069 \pm 0.085	0.00003	0.059 \pm 0.044	0.00001	0.057 \pm 0.037
68E	0.00312	0.077 \pm 0.118	0.00248	0.076 \pm 0.118	0.00003	0.059 \pm 0.045	0.00002	0.058 \pm 0.041
69E	0.00019	0.063 \pm 0.061	0.00043	0.066 \pm 0.072	0.00001	0.068 \pm 0.060	0.00000	0.055 \pm 0.030
70L	0.00302	0.077 \pm 0.117	0.00393	0.081 \pm 0.139	0.00007	0.060 \pm 0.050	0.00015	0.062 \pm 0.056

Table S13. Continued.

Residues	Carnivora		Cetartiodactyla		Chiroptera		Rodentia	
	Probab.	$\omega + SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
71L	0.00483	0.082 + 0.137	0.00684	0.089 \pm 0.178	0.00020	0.062 \pm 0.058	0.00021	0.063 \pm 0.059
72D	0.00029	0.064 + 0.065	0.00043	0.066 \pm 0.072	0.04417	0.408 \pm 0.313	0.00124	0.098 \pm 0.123
73E	0.00456	0.122 + 0.168	0.00043	0.066 \pm 0.072	0.00000	0.054 \pm 0.027	0.00000	0.055 \pm 0.031
74D	0.00027	0.064 + 0.065	0.00043	0.066 \pm 0.072	0.00001	0.068 \pm 0.061	0.00003	0.071 \pm 0.067
75Q	0.00152	0.072 + 0.095	0.00274	0.077 \pm 0.121	0.41732	0.964 \pm 0.483	0.00002	0.059 \pm 0.044
76K	0.00304	0.077 + 0.117	0.09907	0.360 \pm 0.596	0.00208	0.104 \pm 0.135	0.00016	0.062 \pm 0.057
77V	0.20927	0.612 + 0.549	0.00953	0.097 \pm 0.211	0.01180	0.222 \pm 0.236	0.13242	0.557 \pm 0.438
78R	0.00258	0.076 + 0.111	0.00379	0.081 \pm 0.138	0.02526	0.267 \pm 0.285	0.00177	0.099 \pm 0.128
79P	0.00250	0.075 + 0.110	0.00284	0.078 \pm 0.123	0.00127	0.098 \pm 0.122	0.00003	0.059 \pm 0.045
80N *	0.46473	1.038 + 0.626	0.09573	0.353 \pm 0.595	0.14180	0.568 \pm 0.432	0.07289	0.466 \pm 0.362
81E	0.05020	0.235 + 0.360	0.00043	0.066 \pm 0.072	0.00002	0.058 \pm 0.041	0.00003	0.059 \pm 0.044
82E	0.00718	0.132 + 0.189	0.00465	0.083 \pm 0.150	0.00002	0.058 \pm 0.041	0.00001	0.057 \pm 0.038
83N	0.00035	0.065 + 0.068	0.01118	0.150 \pm 0.225	0.00002	0.072 \pm 0.068	0.00148	0.156 \pm 0.163
84N *	0.04190	0.219 + 0.336	0.81383	2.094 \pm 1.218	0.78233	1.346 \pm 0.399	0.88530	1.651 \pm 0.629
85K	0.00024	0.064 + 0.064	0.00058	0.067 \pm 0.077	0.00000	0.055 \pm 0.029	0.00000	0.056 \pm 0.033
86D	0.00027	0.064 + 0.065	0.00046	0.066 \pm 0.073	0.00000	0.054 \pm 0.027	0.00000	0.055 \pm 0.030
87A	0.01098	0.145 + 0.213	0.01938	0.172 \pm 0.277	0.00000	0.056 \pm 0.034	0.00021	0.082 \pm 0.089
88D	0.00027	0.064 + 0.065	0.00043	0.066 \pm 0.072	0.00000	0.054 \pm 0.027	0.00000	0.055 \pm 0.029
89L	0.00554	0.084 + 0.145	0.13986	0.535 \pm 0.568	0.00033	0.064 \pm 0.064	0.00032	0.064 \pm 0.065
90Y	0.00529	0.084 + 0.142	0.00224	0.076 \pm 0.113	0.00003	0.059 \pm 0.044	0.00003	0.059 \pm 0.045
91T	0.00414	0.080 + 0.130	0.00095	0.070 \pm 0.087	0.00729	0.127 \pm 0.183	0.01838	0.160 \pm 0.248
92S	0.00304	0.077 + 0.117	0.00334	0.079 \pm 0.131	0.00010	0.061 \pm 0.052	0.00005	0.060 \pm 0.048
93R	0.00071	0.067 + 0.078	0.00297	0.078 \pm 0.125	0.00000	0.057 \pm 0.036	0.00002	0.058 \pm 0.042
94V	0.00132	0.071 + 0.091	0.00119	0.071 \pm 0.093	0.00000	0.056 \pm 0.034	0.00001	0.057 \pm 0.037
95M	0.00021	0.064 + 0.062	0.00043	0.066 \pm 0.073	0.00000	0.055 \pm 0.028	0.00004	0.075 \pm 0.075
96L	0.00361	0.079 + 0.124	0.00284	0.078 \pm 0.123	0.00004	0.060 \pm 0.047	0.00006	0.060 \pm 0.049

Table S13. Continued.

Residues	Carnivora		Cetartiodactyla		Chiroptera		Rodentia	
	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
97S	0.00278	0.076 + 0.114	0.00068	0.068 ± 0.081	0.00021	0.063 ± 0.059	0.00003	0.059 ± 0.045
98S	0.00319	0.078 + 0.119	0.00648	0.089 ± 0.173	0.00027	0.063 ± 0.062	0.00006	0.060 ± 0.049
99Q	0.00859	0.092 + 0.172	0.01224	0.104 ± 0.241	0.00037	0.064 ± 0.066	0.00097	0.068 ± 0.083
100V	0.00046	0.066 + 0.072	0.00081	0.069 ± 0.084	0.00000	0.056 ± 0.033	0.00001	0.057 ± 0.036
101P	0.00782	0.090 + 0.166	0.01538	0.112 ± 0.275	0.00166	0.071 ± 0.094	0.00213	0.073 ± 0.103
102L	0.00485	0.082 + 0.138	0.00684	0.089 ± 0.178	0.00031	0.064 ± 0.063	0.00032	0.064 ± 0.065
103E	0.00019	0.063 + 0.061	0.00056	0.067 ± 0.076	0.00000	0.055 ± 0.028	0.00000	0.055 ± 0.030
104P	0.00730	0.089 + 0.161	0.00859	0.094 ± 0.200	0.00002	0.059 ± 0.044	0.00043	0.065 ± 0.068
105P	0.00782	0.090 + 0.166	0.01538	0.112 ± 0.275	0.00138	0.070 ± 0.089	0.00033	0.064 ± 0.065
106L	0.00443	0.081 + 0.133	0.00491	0.084 ± 0.153	0.00137	0.070 ± 0.089	0.00040	0.065 ± 0.067
107L	0.00252	0.075 + 0.110	0.00281	0.078 ± 0.123	0.00004	0.060 ± 0.047	0.00003	0.059 ± 0.045
108F	0.00605	0.086 + 0.149	0.00432	0.082 ± 0.145	0.00069	0.066 ± 0.075	0.00073	0.067 ± 0.077
109L	0.00302	0.077 + 0.117	0.00451	0.083 ± 0.148	0.00007	0.060 ± 0.050	0.00012	0.062 ± 0.055
110L	0.00499	0.083 + 0.139	0.00332	0.079 ± 0.130	0.00026	0.063 ± 0.061	0.00052	0.065 ± 0.071
111E	0.00019	0.063 + 0.061	0.00043	0.066 ± 0.072	0.00000	0.054 ± 0.027	0.00000	0.055 ± 0.030
112E	0.00312	0.077 + 0.118	0.00465	0.083 ± 0.150	0.00003	0.059 ± 0.045	0.00004	0.059 ± 0.046
113Y	0.00177	0.073 + 0.099	0.00197	0.075 ± 0.108	0.00003	0.059 ± 0.044	0.00002	0.059 ± 0.044
114K	0.00397	0.080 + 0.128	0.00534	0.085 ± 0.159	0.00008	0.061 ± 0.051	0.00031	0.064 ± 0.065
115N	0.00173	0.073 + 0.099	0.00461	0.083 ± 0.149	0.00006	0.060 ± 0.049	0.00015	0.062 ± 0.056
116Y	0.00177	0.073 + 0.099	0.00197	0.075 ± 0.108	0.00005	0.060 ± 0.047	0.00002	0.059 ± 0.044
117L	0.00302	0.077 + 0.117	0.00451	0.083 ± 0.148	0.00009	0.061 ± 0.052	0.00012	0.062 ± 0.055
118D	0.00138	0.071 + 0.093	0.00327	0.079 ± 0.130	0.00002	0.059 ± 0.043	0.00001	0.057 ± 0.038
119A	0.00401	0.080 + 0.129	0.00699	0.090 ± 0.180	0.00046	0.065 ± 0.068	0.00015	0.062 ± 0.055
120A	0.00829	0.091 + 0.170	0.00188	0.074 ± 0.107	0.00037	0.064 ± 0.065	0.00028	0.063 ± 0.062
121N	0.00024	0.065 ± 0.069	0.00045	0.066 ± 0.073	0.00000	0.055 ± 0.030	0.00000	0.056 ± 0.032
122M			0.00043	0.066 ± 0.073	0.00000	0.055 ± 0.028	0.00000	0.056 ± 0.032

Table S13. Continued.

Residues	Carnivora		Cetartiodactyla		Chiroptera		Rodentia	
	Probab.	$\omega + SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
123S	0.00761	0.089 + 0.164	0.00327	0.079 \pm 0.130	0.00050	0.065 \pm 0.070	0.00022	0.063 \pm 0.060
124M	0.00510	0.124 + 0.173	0.00043	0.066 \pm 0.073	0.00000	0.055 \pm 0.028	0.00000	0.056 \pm 0.032
125R ⁴	0.00052	0.067 + 0.074	0.00065	0.068 \pm 0.080	0.00000	0.057 \pm 0.036	0.00005	0.060 \pm 0.048
126V ⁴	0.00086	0.069 + 0.082	0.00102	0.070 \pm 0.089	0.00001	0.057 \pm 0.037	0.00001	0.057 \pm 0.037
127R ⁴	0.00209	0.074 + 0.104	0.00259	0.077 \pm 0.119	0.00007	0.060 \pm 0.049	0.00009	0.061 \pm 0.053
128R ⁴	0.00245	0.075 + 0.109	0.00241	0.076 \pm 0.116	0.00006	0.060 \pm 0.049	0.00004	0.060 \pm 0.047
129H	0.00100	0.069 + 0.085	0.00140	0.072 \pm 0.098	0.00000	0.057 \pm 0.036	0.00001	0.057 \pm 0.037
130S	0.00778	0.090 + 0.165	0.00348	0.080 \pm 0.133	0.00031	0.064 \pm 0.063	0.00009	0.061 \pm 0.052
131D	0.00027	0.064 + 0.065	0.00043	0.066 \pm 0.072	0.00000	0.054 \pm 0.027	0.00000	0.055 \pm 0.029
132P	0.00291	0.077 + 0.116	0.00389	0.081 \pm 0.139	0.00003	0.059 \pm 0.045	0.00093	0.068 \pm 0.081
133A	0.01893	0.167 + 0.253	0.00100	0.070 \pm 0.089	0.00001	0.057 \pm 0.037	0.00001	0.057 \pm 0.038
134R	0.00245	0.075 + 0.109	0.00241	0.076 \pm 0.116	0.00006	0.060 \pm 0.049	0.00004	0.060 \pm 0.047
135R	0.00675	0.087 + 0.156	0.00245	0.076 \pm 0.117	0.00007	0.060 \pm 0.050	0.00103	0.069 \pm 0.084
136G	0.00038	0.065 + 0.069	0.00056	0.067 \pm 0.077	0.00000	0.054 \pm 0.027	0.00001	0.057 \pm 0.036
137E	0.00019	0.063 + 0.061	0.00043	0.066 \pm 0.072	0.00007	0.060 \pm 0.050	0.00000	0.055 \pm 0.030
138L	0.00302	0.077 + 0.117	0.00450	0.083 \pm 0.148	0.00000	0.057 \pm 0.036	0.00013	0.062 \pm 0.055
139S	0.00074	0.068 + 0.080	0.00068	0.068 \pm 0.081	0.00000	0.056 \pm 0.033	0.00001	0.057 \pm 0.038
140V	0.00048	0.066 + 0.072	0.00081	0.069 \pm 0.084	0.00019	0.062 \pm 0.058	0.00001	0.057 \pm 0.036
141C	0.00286	0.077 + 0.115	0.00283	0.078 \pm 0.123	0.00000	0.054 \pm 0.027	0.00043	0.065 \pm 0.068
142D	0.00027	0.064 + 0.065	0.00043	0.066 \pm 0.072	0.00000	0.057 \pm 0.036	0.00000	0.055 \pm 0.029
143S	0.01636	0.161 + 0.242	0.00068	0.068 \pm 0.081	0.00203	0.102 \pm 0.133	0.00007	0.060 \pm 0.050
144I ⁵	0.00299	0.077 + 0.117	0.00099	0.070 \pm 0.088	0.00000	0.057 \pm 0.036	0.01336	0.229 \pm 0.244
145S	0.00075	0.068 + 0.080	0.00074	0.069 \pm 0.082	0.00000	0.054 \pm 0.027	0.00001	0.057 \pm 0.039
146E ⁵	0.00019	0.063 + 0.061	0.00043	0.066 \pm 0.072	0.00026	0.063 \pm 0.062	0.00000	0.055 \pm 0.030
147W	0.00461	0.082 + 0.135	0.00535	0.085 \pm 0.159	0.00026	0.063 \pm 0.062	0.00027	0.064 \pm 0.063

Table S13. Continued.

Residues	Carnivora		Cetartiodactyla		Chiroptera		Rodentia	
	Probab.	$\omega + SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
148V	0.00046	0.066 + 0.072	0.00081	0.069 \pm 0.084	0.00005	0.060 \pm 0.048	0.00002	0.058 \pm 0.040
149T	0.00805	0.090 + 0.168	0.00098	0.070 \pm 0.088	0.00015	0.062 \pm 0.056	0.00006	0.060 \pm 0.049
150A	0.00049	0.066 + 0.073	0.00106	0.070 \pm 0.090	0.00001	0.058 \pm 0.041	0.00001	0.057 \pm 0.039
151A	0.00111	0.070 + 0.088	0.00088	0.069 \pm 0.086	0.00006	0.060 \pm 0.048	0.00051	0.065 \pm 0.070
152D	0.00120	0.070 + 0.089	0.00399	0.081 \pm 0.140	0.00002	0.058 \pm 0.041	0.00005	0.060 \pm 0.047
153K	0.00397	0.080 + 0.128	0.00423	0.082 \pm 0.145	0.00005	0.060 \pm 0.048	0.00031	0.064 \pm 0.065
154K⁶	0.00023	0.064 + 0.063	0.00044	0.066 \pm 0.073	0.00000	0.055 \pm 0.029	0.00000	0.056 \pm 0.033
155T⁶	0.00481	0.082 + 0.137	0.00893	0.095 \pm 0.204	0.00066	0.066 \pm 0.074	0.00050	0.065 \pm 0.070
156A⁶	0.00829	0.091 + 0.170	0.00747	0.091 \pm 0.188	0.00025	0.063 \pm 0.060	0.00060	0.066 \pm 0.073
157V⁶	0.00046	0.066 + 0.072	0.00081	0.069 \pm 0.084	0.00000	0.056 \pm 0.033	0.00001	0.057 \pm 0.036
158D⁶	0.00027	0.064 + 0.065	0.00043	0.066 \pm 0.072	0.00000	0.054 \pm 0.027	0.00000	0.055 \pm 0.029
159M⁶	0.00021	0.064 + 0.062	0.00043	0.066 \pm 0.073	0.00000	0.055 \pm 0.028	0.00000	0.056 \pm 0.032
160S⁶	0.00296	0.077 + 0.116	0.00594	0.087 \pm 0.167	0.00013	0.062 \pm 0.055	0.00019	0.063 \pm 0.059
161G⁶	0.00073	0.068 + 0.079	0.00087	0.069 \pm 0.085	0.00000	0.057 \pm 0.036	0.00013	0.062 \pm 0.055
162G⁶	0.00038	0.065 + 0.069	0.00059	0.067 \pm 0.078	0.00000	0.056 \pm 0.033	0.00001	0.057 \pm 0.038
163T⁶	0.00063	0.067 + 0.077	0.00206	0.075 \pm 0.110	0.00000	0.057 \pm 0.037	0.00007	0.060 \pm 0.050
164V	0.00086	0.069 + 0.082	0.00110	0.070 \pm 0.091	0.00001	0.057 \pm 0.037	0.00001	0.057 \pm 0.036
165T	0.00091	0.069 + 0.083	0.00122	0.071 \pm 0.094	0.00001	0.057 \pm 0.037	0.00121	0.069 \pm 0.087
166V	0.00086	0.069 + 0.082	0.00102	0.070 \pm 0.089	0.00001	0.057 \pm 0.037	0.00001	0.057 \pm 0.037
167L	0.00671	0.087 + 0.156	0.00405	0.082 \pm 0.141	0.00058	0.066 \pm 0.072	0.00013	0.062 \pm 0.055
168E	0.00312	0.077 + 0.118	0.00465	0.083 \pm 0.150	0.00003	0.059 \pm 0.044	0.00003	0.059 \pm 0.043
169K	0.00385	0.080 + 0.127	0.00491	0.084 \pm 0.153	0.00003	0.059 \pm 0.045	0.00002	0.058 \pm 0.042
170V	0.00086	0.069 + 0.082	0.00102	0.070 \pm 0.089	0.00001	0.057 \pm 0.037	0.00001	0.057 \pm 0.036
171P	0.00401	0.080 + 0.129	0.00291	0.078 \pm 0.124	0.00096	0.068 \pm 0.081	0.00007	0.060 \pm 0.050
172V	0.00611	0.086 + 0.150	0.00102	0.070 \pm 0.089	0.00009	0.061 \pm 0.051	0.00008	0.060 \pm 0.050
173S⁶			0.00355	0.080 \pm 0.134	0.00011	0.061 \pm 0.053	0.00011	0.061 \pm 0.054

Table S13. Continued.

Residues	Carnivora		Cetartiodactyla		Chiroptera		Rodentia	
	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$	Probab.	$\omega \pm SE$
174K ⁶	0.00397	0.080 + 0.128	0.00534	0.085 ± 0.159	0.00008	0.061 ± 0.051	0.00031	0.064 ± 0.065
175G ⁶	0.00073	0.068 + 0.079	0.00072	0.068 ± 0.082	0.00000	0.057 ± 0.036	0.00000	0.057 ± 0.036
176K ⁶	0.00460	0.082 + 0.135	0.00674	0.089 ± 0.179	0.00022	0.063 ± 0.059	0.00030	0.064 ± 0.064
177L ⁶	0.00302	0.077 + 0.117	0.00451	0.083 ± 0.148	0.00007	0.060 ± 0.050	0.00012	0.062 ± 0.055
178K	0.00023	0.064 + 0.063	0.00044	0.066 ± 0.073	0.00000	0.055 ± 0.029	0.00000	0.056 ± 0.033
179Q	0.00152	0.072 + 0.095	0.00274	0.077 ± 0.121	0.00001	0.057 ± 0.039	0.00065	0.066 ± 0.075
180Y	0.00177	0.073 + 0.099	0.00197	0.075 ± 0.108	0.00003	0.059 ± 0.044	0.00004	0.060 ± 0.047
181F	0.00201	0.074 + 0.103	0.00249	0.077 ± 0.117	0.00003	0.059 ± 0.046	0.00002	0.059 ± 0.044
182Y	0.00438	0.081 + 0.133	0.00197	0.075 ± 0.108	0.00003	0.059 ± 0.044	0.00002	0.059 ± 0.044
183E	0.00019	0.063 + 0.061	0.00043	0.066 ± 0.072	0.00000	0.054 ± 0.027	0.00000	0.055 ± 0.030
184T	0.00120	0.070 + 0.090	0.00116	0.071 ± 0.092	0.00001	0.058 ± 0.040	0.00001	0.057 ± 0.039
185K	0.00024	0.064 + 0.063	0.00044	0.066 ± 0.073	0.00000	0.055 ± 0.030	0.00000	0.056 ± 0.034
186C	0.00284	0.077 + 0.115	0.00264	0.077 ± 0.120	0.00010	0.061 ± 0.053	0.00010	0.061 ± 0.053
187N	0.00173	0.073 + 0.099	0.00461	0.083 ± 0.149	0.00003	0.059 ± 0.045	0.00015	0.062 ± 0.056
188P	0.00256	0.076 + 0.111	0.00321	0.079 ± 0.129	0.00123	0.098 ± 0.122	0.00004	0.059 ± 0.046
189M		-	0.00043	0.066 ± 0.073	0.00000	0.055 ± 0.028	0.00000	0.056 ± 0.032
190G		-	0.00714	0.090 ± 0.183		-	0.00003	0.059 ± 0.044
191Y		-	0.00197	0.075 ± 0.108		-		-
192T		-	0.00807	0.093 ± 0.194		-		-
193K		-	0.00044	0.066 ± 0.073		-		-
194E		-	0.00043	0.066 ± 0.072		-		-
195G		-	0.00072	0.068 ± 0.082		-		-
196C		-	0.00264	0.077 ± 0.120		-		-
197R		-	0.00107	0.070 ± 0.089		-		-
198G		-		-		-		-
199I		-		-		-		-

Supplementary Figures.

```
MTILFLTMVISYFGCMKAAPMKEANIRGQGLAYPGVIRTHGTLESVNGPKAGSRGLTSLADTFEHVIEEL 70
LDEDQKVRPNEENNKDADLYTSRVMLSSQVPLEPPLLFLLEEYKNYLDAANMSMRVRRHSDPARRGELSV 140
CDSISEWVTAADKKTAVDMSGGTVTVLEKVPVSKGQLKQYFYETKCNPMGYTKEGCRGIDKRHWNSQCRT 210
TQSYVRALTMDSKKRIGWRFIRIDTSCVCTLTIKRGR
```

Fig. S1. Human BDNF *isoform a*, that codes for a protein containing 247 amino acids. The first 18 residues comprise the signal peptide. The underlined motifs represent sites within BDNF that have been implicated in functional aspects: cleavage to form a shorter proBDNF (54RGLT↓SL59) (Seidah *et al.*, 1999); mature form cleavage (125RVRR128) (Lee *et al.*, 2001; Gray & Ellis, 2008); carboxipeptidase-E co-receptor interaction (144I and 146E; 233ID234) (Luo *et al.*, 2005); TrkB receptor interaction (154KTAVDMSGGT163; 173SKGQL177; 207QCRTTQSYVR216) (Ibáñez *et al.*, 1991; Ibáñez *et al.*, 1993); interaction with receptors p75 and TrkB (223KKR225) (Rydén *et al.*, 1995).



Figure S2. Alignment of the target-sequence (human locus *BDNF* - p23560) - excluding the signal peptide - and the two templates, which have 3D structures available in the protein data bank (pdb): 1b8m subunit A and 2ia8. 2ia8 subunit A, derived from mitochondrial cytochrome c peroxidase, was identified as the best available match using BLAST screening, with 30% identities. 1b8m subunit A was obtained directly from BDNF crystallographic studies (for more information see Supplementary Methods). The figure was generated using the multiple sequence alignment program T-Coffee. Symbols: “*” indicates regions of amino acid identities; “:” indicates regions of amino acid similarities; “.” indicates regions of unrelated amino acids.

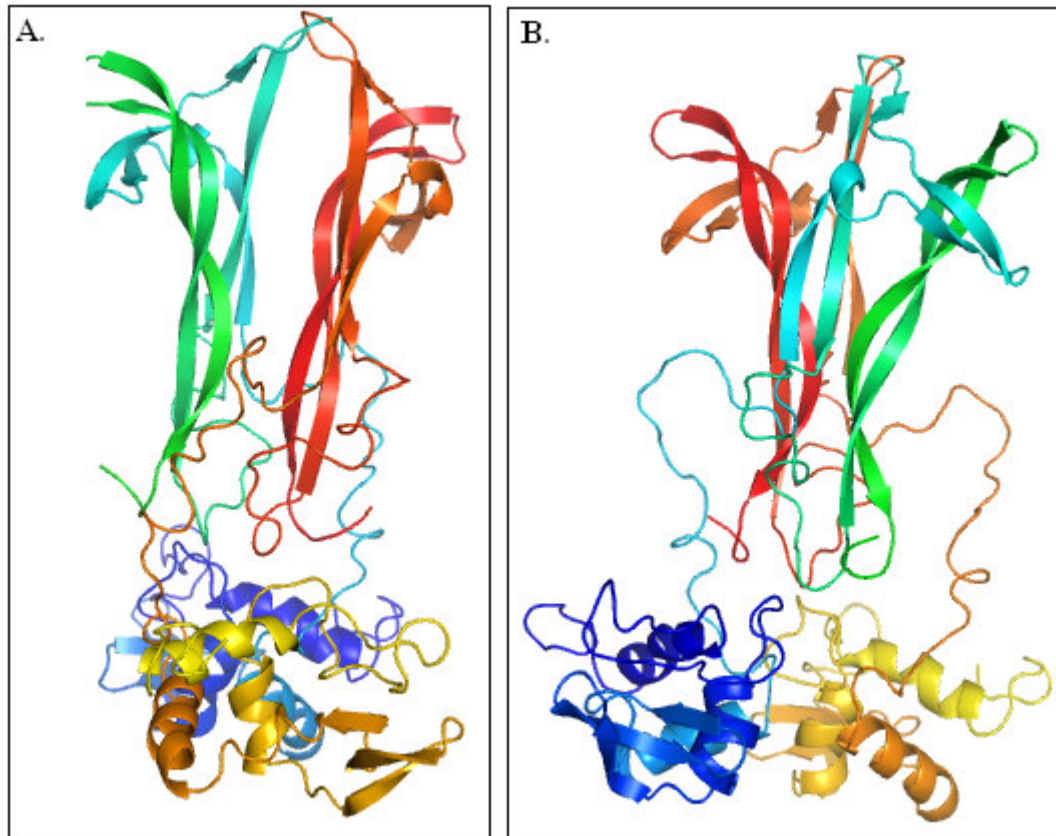


Figure S3. ProBDNF homodimer, with two symmetric units of the mature domain (green and red colors), and two symmetric units of the pro domain (blue and yellow colors). The figures were generated using the program Pymol (version 0.99rc6). **A** and **B** represent distinct views.



Fig. S4. Location of positively selected amino acids (underlined) identified in each data set, using one representative sequence (without the signal peptide) as an example. Grey fonts indicate the pro domain, whereas black fonts depict the mature domain. **A.** Protein sequence of *Homo sapiens* (data set 1); **B.** Protein sequence of *Homo sapiens* (data set 2); **C.** Protein sequence of cattle (*Bos taurus*, order Cetartiodactyla); the rectangle indicates an insertion of three serines in this sequence; **D.** Protein sequence of a mountain paca (*Agouti taczanowskii*, order Rodentia). The rectangle indicates an insertion of 11 serines observed in this sequence; **E.** Protein sequence of sheath-tailed bat (*Emballonura atrata*, order Chiroptera); **F.** Protein sequence of a domestic cat (*Felis catus*) (order Carnivora).

Capítulo 3 – Considerações Finais.

Os estudos conjuntos de análises evolutivas de seqüências e análises estruturais das mesmas proteínas (enfocando conformações em momentos distintos) e suas implicações com características funcionais podem contribuir para elucidar aspectos funcionais desconhecidos, entender em maiores detalhes o processo de evolução da molécula em questão, e até mesmo redirecionar pesquisas experimentais da área envolvida.

Esta investigação indicou regiões do BDNF conservadas entre espécies, assim como o padrão de seleção natural atuante (seleção positiva, negativa ou neutra) em cada um de seus codons. Estas informações indicaram que a restrição seletiva negativa, atuando em uma proporção similar em toda seqüência, pôde ocorrer no curso da evolução das linhagens de mamíferos. Esta análise foi administrada em uma caracterização estrutural paralela abordando a construção de um modelo viável condizente com a predição da estrutura secundária e a conformação nos domínios do BDNF (pro e maduro), baseado na modelagem molecular por homologia do peptídeo inteiro.

Apesar de o domínio maduro do BDNF formar cristais heterodiméricos estáveis com outros membros da família (BDNF/NT3; BDNF/NT4), ainda não havia um modelo de duas unidades simétricas da porção madura desta proteína disponível, modelo este importante para inferências estruturais, em nível atômico. Recentes estudos de *Paoletti et al. (2006)* e *Paoletti et al. (2008)* demonstraram que o domínio pró não processado do NGF (membro mais conhecido desta família protéica) forma também uma estrutura dimérica simétrica. Estes estudos apontaram dois esboços estruturais para o peptídeo inteiro, baseados nas análises experimentais com SAXS. Nossos modelos do proBDNF inteiro foram condizentes

com o formato *crab-like*, descrito por estes autores, na qual um monômero é constituído por um domínio pró, ancorado pela “base” de um domínio maduro. Os modelos gerados demonstraram-se satisfatórios na maioria dos parâmetros inferidos. A estrutura peculiar de todo proBDNF remeteu-nos a uma maior cautela na modelagem, assim como a exclusão de modelos em muitas tentativas anteriores.

Estudos experimentais que indicaram motivos consenso relacionados à função do BDNF nos tecidos-alvos identificaram, até o momento, apenas motivos situados no domínio maduro. Através das nossas análises, tanto o domínio pró quanto o domínio maduro demonstraram uma proporção similar de restrição seletiva negativa. Levando em consideração aqueles estudos anteriores e o nosso, infere-se uma maior atenção funcional e estrutural para o domínio pró, do qual se conhece pouco, como também sua relação com o domínio maduro. Neste contexto, a função do BDNF foi identificada a partir de estudos comparativos, utilizando o NGF, e posteriormente confirmado por métodos bioquímicos e de localização celular. A partir daí, as pesquisas em neurociências têm apostado nesta pequena proteína, dentre outros aspectos, por causa de suas particularidades importantes, que claramente o diferenciam dos outros membros neurotróficos, incluindo plasticidade neuronal (relevante para o aprendizado e a memória), e um polimorfismo funcional bem relatado na literatura (V66M). Entretanto, questões ainda teóricas - se o pró-domínio é ou não um intermediário na rota de sinalização, se é rapidamente clivado para a proteína madura, e restrições práticas para o estudo *in vitro* utilizando o pró-domínio, têm dificultado o direcionamento do foco das pesquisas para a região pró.

O polimorfismo humano V66M, amplamente citado na literatura, implica uma substituição única do aminoácido valina por uma metionina no domínio pró. Além de algumas patologias envolvidas, este SNP afeta o direcionamento do domínio maduro processado para uma rota de sinalização regulada. Outro intrigante fato descrito diz

respeito à inserção dos aminoácidos serina e/ou treonina na seqüência pró de algumas espécies (46 de 153 analisadas), primeiramente observada para poucas espécies, e ampliada pelo nosso estudo. Nosso estudo indentificou um padrão de variação de inserção de uma até onze serinas. Em roedores utilizados como modelo para os estudos acerca do BDNF (*Mus musculus*, *Rattus norvegicus*) foi observado à inserção de uma treonina seguida pela serina. Através do parâmetro de ômega (ω) analisado, e a posição destas inserções na estrutura modelada, sugere-se que estes *indels* estão próximos de sítios sujeitos a restrições seletivas importantes, o que indica que a presença destes aminoácidos adicionais pode ter efeitos biológicos relevantes. A contribuição de futuras pesquisas que adicionem estas inserções seria relavante para relacionar com possíveis implicações funcionais. Em geral, esta pesquisa encoraja futuras contribuições, que além de outros aspectos, almejem elucidar e aprofundar os estudos iniciais de interações do BDNF com outras moléculas, usando como base os sítios identificados neste estudo como provavelmente exercendo funções biológicas importantes.

Referências Bibliográficas.

- Aid, T.; Kazantseva, A.; Piirsoo, M. et al. Mouse and rat BDNF gene structure and expression revisited. *J. Neurosci. Res.*, v. 85, pp. 525–35, 2007.
- Ammirante, M.; Rosati, A.; Gentilella, A. et al. The activity of hsp90 alpha promoter is regulated by NF-kappa B transcription factors. *Nature*, v. 27, n.8, pp. 1175-8, 2007.
- Arab, S.; Krohn, K.; Lachmund, A. et al. The gene encoding bovine brain-derived neurotrophic factor (BDNF). *Gene*, v.185, n.1, pp. 95-8, 1997.
- Balasubramanian, S.; Zheng, D.; Liu, Y. et al. Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes, *Genome Biol.*, v.10, n.1, pp. R2, 2009.
- Barde, Y.A; Edgar, D; Thoenen, H. Purification of a new neurotrophic factor from mammalian brain. *EMBO J.*, v.1, n.5, pp 549-53, 1982.
- Berman, H.; Westbrook, J.; Feng, Z. et al. The Protein Data Bank. *Nucl. Ac. Res.*, v.28, pp. 235- 242, 2000.
- Bernadó, P.; Mylonas, E.; Petoukhov, M. et al. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*, v.129, n. 17, pp. 5656-64, 2007.
- Binder, D.; Croll, S.; Gall, C. et al. BDNF and epilepsy: too much of a good thing?. *Trends in Neurosci.*, v. 24, pp. 47-53, 2001.
- Bolamos, C.; Nestler, E. Neurotrophic mechanisms in drug addiction. *NeuroMol. Med.*, v. 5, pp. 69-83, 2004.
- Boutillier, J.; Ceni, C.; Pagdala, P. et al. Proneurotrophins require endocytosis and intracellular proteolysis to induce TrkA activation. *JBC*, v. 263, n. 19, pp. 12709 – 16, 2008.
- Butte, M.; Hwang, P.; Mobley, W.; Fletterick, R. Crystal structure of neurotrophin-3 homodimer shows distinct regions are used to bind its receptors. *Biochem*, v. 37, pp. 16846-52, 1998.
- Butte, M. Neurotrophic factor structures reveal clues to evolution, binding, specificity, and receptor activation. *Cell Mol Life Sci*, v. 58, pp. 1003-13, 2001.
- Chen, Z.; Jing, D.; Bath, K. et al. Genetic Variant BDNF (Val66Met) Polymorphism Alters Anxiety-Related Behavior, *Science*, v. 314, pp. 140 – 43, 2006.
- Colovos, C; Yeates, TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.*, v. 2, n. 9, pp.1511-9, 1993.
- Da Silva, V; Silva, C. Modelagem molecular de proteínas-alvo por homologia estrutural. *REF*, v.4, n.1, PP. 15-26, 2007.
- Deane, C.; Blundell, T. Protein comparative modeling and drug discovery. In: WERMUTH, C. The practice of medical chemistry. London: *Elsevier Academic Press*, 2003.
- DeLano, W. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA, 2002. [<http://www.pymol.org>].
- Egan, M.; Kojima, M.; Callicott, J. et al. The BDNF val66met Polymorphism Affects Activity-Dependent Secretion of BDNF and Human Memory and Hippocampal Function. *Cell*, v. 112, pp.257–69, 2003.
- Eick, G.; Jacobs, D.; Matthee, C. A nuclear DNA phylogenetic perspective on the evolution of echolocation and historical biogeography of extant bats (Chiroptera). *Mol Biol Evol.*, v. 22, pp. 1869–1886, 2005.

- Ferrer, I.; Marín, C.; Rey, M. Et al. BDNF and full-length and truncated TrkB expression in Alzheimer disease. Implications in therapeutic strategies. *J. Neuropathol Exp Neurol.*, v.58, p.729-39, 1999.
- Ewens, W.; Grant, G. BLAST. In: *Statistical methods in bioinformatics: An introduction (Statistic for biology and health)*. Hardcover: Springer, 1ed, 2001.
- Givnish T.; Sytsma, K. *Molecular Evolution and Adaptive Radiation*. Cambridge: Cambridge University Press; 1997.
- Gray, K.; Ellis, V. Activation of pro-BDNF by the pericellular serine protease plasmin. *FEBS Lett.*, v.582, pp. 907–10, 2008.
- Guex, N.; Peitsch, M. SWISS-MODEL and The Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, v. 18, pp. 2714 – 23, 1997.
- Hallböök, F. Evolution of the vertebrate neurotrophin and Trk receptor gene families. *Science*, v. 9, pp. 616-21, 1999.
- Hallböök, F.; Lundin, L.; Kullander, K. Lampetra fluviatilis Neurotrophin Homolog, Descendant of a Neurotrophin Ancestor, Discloses the Early Molecular Evolution of Neurotrophins in the Vertebrate Subphylum. *J. Neurosci.*, v.18, n.21, pp. 8700-11, 1998.
- Höltje, H.; Sippl, W.; Rognan, D.; Folkers, G. Introduction to comparative protein modeling. In: *Molecular modeling: Basic principles and applications*. Weinheim: Wiley-VCH, 2003.
- Hong, C.; Huo, S.; Yen, F. et al. Association study of a brain-derived neurotrophic-factor genetic polymorphism and mood disorders, age of onset and suicidal behavior. *Neuropsychobiol.*, v.48, pp.186-89, 2003.
- Howells, D.; Porritt, M.; Wong, J. et al. Reduced BDNF mRNA Expression in the Parkinson's Disease Substantia Nigra. *Exp. Neuro.*, v.166, pp. 127-35, 2000.
- Jiang, X.; Xu, K.; Hoberman, J. et al. BDNF variation and mood disorders: a novel functional promoter polymorphism and Val66Met are associated with anxiety but have opposing effects. *Neuropsychopharmacolog.*, v.30, pp.1353-61, 2005.
- Junkbluth, S; Bailey, K; Barde, Y-A. Purification and characterisation of brain-derived neurotrophic factor / neurotrophin-3 (BDNF / NT-3) heterodimer. *Eur J Biochem*, v. 221, n. 2, pp. 677 – 85, 2005.
- Katoh-Semba, R; Takeuchi, I.; Semba, R.; Kato, K. Distribution of brain-derived neurotrophic factor in rats and its changes with development in the brain. *J. Neurochem.*, v. 69, pp. 34–42, 1997.
- Kawai K.; Nikaido M.; Harada M. et al. Intra and interfamily relationships of Vespertilionidae inferred by various molecular markers including SINE insertion data. *J Mol Evol.* , v. 55, pp. 284–301. 2002.
- Kliemann, M.; Golbik, R.; Rudolph, R. et al. The pro-peptide of pro-NGF: Structure formation and intramolecular association with NGF. *Protein Science*, v. 16, pp. 411 – 19, 2007.
- Kliemann, M.; Rathenhol, A.; Golbik. et al. The mature part of proNGF induces the structure of its pro-peptide. *FEBS Letters*, v. 566, pp. 207 – 12, 2004.
- Koepfli, K.;Deere, K.; Slater, G. et al. Multigene phylogeny of the Mustelidae: resolving relationships, tempo and biogeographic history of a mammalian adaptive radiation. *BMC Biol.*, v. 14, pp. 6-10, 2008.
- Korneev, S.; O'Shea, M. Natural antisense RNAs in the nervous system. *Rev. Neurosci*, v. 16, pp. 213-22, 2005.
- Koshimizu, H.; Kiyosue, K.; Hara, T., et al. Multiple functions of precursor BDNF to CNS neurons: negative regulation of neurite growth, spine formation and cell survival. *Mol brai*, pp. 2-27, 2009.

- Kovalchuk, Y.; Hanse, E.; Kafitz, K.; Konnerth, A. Postsynaptic Induction of BDNF-mediated Long-Term Potentiation. *Science*, v.295, pp 1729-34, 2002.
- Kullander, K.; Carlson, B.; Hallböök, F. Molecular phylogeny and evolution of the neurotrophins from monotremes and marsupials. *J. Mol. Evol.*, v. 45, pp. 311–321, 1997.
- Lanave, C.; Colangelo, A.; Saccone, C.; Alberghina, L. Molecular evolution of the neurotrophin family members and their Trk receptors. *Gene*, v.394, pp. 1-12, 2007.
- Lanktree, M.; Squassina, A.; Krinsky, M. et al. Association Study of Brain-Derived Neurotrophic Factor (BDNF) and LIN-7 Homolog (LIN-7) Genes with Adult Attention-Deficit/Hyperactivity Disorder. *Am J Med Genet B Neuropsychiatr Genet.*, v.147, n. 6, pp. 945-51, 2008.
- Lebmann, V.; Brigadski, T. Mechanisms, locations, and kinetics of synaptic BDNF secretion: An update. *Neurosci Res*, v. 65, pp. 11–22, 2009.
- Lee, R.; Kerami, P.; Teng, K.; Hempstead, B. Regulation of cell survival by secreted proneurotrophins. *Science*, v. 294, pp.1945–48, 2001.
- Leibrock, J.; Lottspeich, F.; Hohn, A. et al. Molecular cloning and expression of brain derived neurotrophic factor. *Nature*, v.341, n.6238, pp.149-52, 1989.
- Lesk, A. *Introdução à bioinformática*. Porto Alegre: Artmed, 2 ed, 2008.
- Lou, H.; Kim, S-K; Zaitsev, E. et al. Sorting and Activity-Dependent Secretion of BDNF Require Interaction of a Specific Motif with the Sorting Receptor Carboxypeptidase E. *Neuron*, v. 45, n. 2, pp. 245-55, 2005.
- Lüthy, R.; Bowie, J.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature*, v. 356, pp. 83 -85, 1992.
- Martí-Renom, M.; Stuart, A.; Fiser, A. et al. Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct*, v. 29, pp. 291 – 325. 2000.
- Matsumoto, T.; Rauskolb, S.; Polack, M. et al. Biosynthesis and processing of endogenous BDNF: CNS neurons store and secrete BDNF, not pro-BDNF. *Nat Neurosci.*, v. 11, n. 2, pp. 131-33, 2008.
- Mercader, J.; Ribasés, M.; Gratacòs, M. et al. Altered brain-derived neurotrophic factor blood levels and gene variability are associated with anorexia and bulimia. *Genes, brain and behavior*, v.6, pp. 706-16, 2007.
- Miller-Butterworth, C.; Murphy, W.; O'Brien, S. et al. A Family Matter: Conclusive Resolution of the Taxonomic Position of the Long-fingered Bats, *Mniopterus*, *Mol. Biol. Evol.*, v. 24, n.7, pp.1553-61, 2007.
- Mowla, S.; Farhadi, H.; Pareek, S. et al. Biosynthesis and post-translational processing of the precursor to Brain-Derived Neurotrophic Factor. *J. Biol. Chem.*, v. 276, n. 16, pp. 12660-66, 2001.
- Murphy, W.; Eizirik, E.; O'Brien, S. et al. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*, v. 294, n. 5550, pp. 2348-51, 2001.
- Nagappan, G.; Zaitsev, E.; Senatorov, V. et al. Control of extracellular cleavage of proBDNF by high frequency neuronal activity. *Proc. Natl. Acad. Sci. USA*, v. 106, n.4, pp. 1267-72, 2009.
- Nayeem, A.; Sitkoff, D.; Junior, S. A comparative study of available software for high accuracy homology modeling: from sequence alignments to structural models. *Protein Sci.*, v.15, pp.808-24, 2006.
- Nomoto, H.; Takaiwa, M.; Mouri, A.; Furukawa, S. Pro-region of neurotrophins determines the processing efficiency. *Biochem Biophys Res Commun*, v. 356, n.4, pp. 919 – 24, 2007.
- Paiardini, A.; Caputa, V. Insights into the interaction of sortilin with proneurotrophins: A computational approach. *Neuropeptides*, v. 42, pp. 205 – 14, 2008.

- Pang, P.; Teng, H.; Zaitsev, E. et al. Cleavage of proBDNF by tPA/Plasmin is essential for long-term hippocampal plasticity. *Science*, v. 306, n. 5695, pp. 487- 91, 2004.
- Paoletti, F.; Covaceuszach, S.; Konarev, P. et al. Intrinsic structural disorder of mouse proNGF. *Proteins*, v. 75, pp. 990 – 1009, 2009.
- Park, A.; Holmer, L.; Worman, H. A human HP1 pseudogene maps to chromosome 11p14. *Somat Cell Mol Genet.*, v.24, n.6, pp. 353-6, 1998.
- Pearlman, D.; Case, D; Caldwell, J. et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Commun*, v. 91, pp. 1-41, 1995.
- Pillai, A. Brain-Derived Neurotrophic Factor/TrkB Signaling in the Pathogenesis and Novel Pharmacotherapy of Schizophrenia, *Neurosignals*, v.16, pp. 183–93, 2008.
- Poo, M. Neurotrophins as synaptic modulators. *Nature*, v.2, pp. 24-32, 2001.
- Pruunsild, P.; Kazantseva, A.; Aid, T. et al. Dissecting the human BDNF locus: bidirectional transcription, complex splicing, and multiple promoters. *Genomics*, v. 90, n.3, pp.397-406, 2007.
- Rattenholl, A.; Lilie, H.; Grossmann, A. et al. The pro-sequence facilitates folding of human Nerve Growth Factor from *Escherichia coli* inclusion bodies. *Eur. J. Biochem.*, v. 268, pp. 3296 -3303, 2001b.
- Rattenholl, A.; Ruoppolo, M.; Flagiello, A. et al. Pro-sequence assisted folding and disulfide bond formation of human Nerve Growth Factor. *J. Mol. Biol.*, v. 305, pp. 523 – 33, 2001a.
- Reichardt, L. Neurotrophin-regulated signalling pathways. *Phil. Trans. R. Soc. B*, v.361, pp. 1545–64, 2006.
- Robinson, R.; Radziejewski, C.; Stuart, D.; Jones, E. Structure of the Brain-Derived Neurotrophic factor/ Neurotrophin 3 Heterodimer. *Biochem.*, v. 34, pp. 4139-45, 1995.
- Robinson, R.; Radziejewski, C.; Spraggon, G; et al. The structures of the neurotrophin 4 homodimer and the brain-derived neurotrophic factor/neurotrophin 4 heterodimer reveal a common Trk-binding site. *Protein Sci.*, v.8, pp. 2589-97, 1999.
- Sali, A.; Blundell, T. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol*, v.234, n.3, pp. 779 – 815, 1993.
- Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. Swiss-Model: an automated protein homology-moleling server. *Nucl Ac Res*, v. 31, n. 13, pp. 3381-85, 2003.
- Seidah, N.; Mowla, S.; Hamelin, J. et al. Mammalian subtilisin / Kexin isozyme SKI-1: A widely expressed proprotein convertase with a unique cleavage specificity and cellular localization. *Proc. Natl. Acad. Sci. USA*, v. 96, pp. 1321 – 26, 1999.
- Teng, H.; Teng, K.; Lee, R. et al. ProBDNF induces neuronal apoptosis via activation of a receptor complex of p75NTR and sortilin, *J. Neurosci.*, v. 25, pp.5455–63, 2005.
- Van Den Bussche, R.; Hofer, S. Phylogenetic relationships among recent chiropteran families and the importance of choosing appropriate out-group taxa. *J Mammal*, v. 85, pp. 321–30, 2004.
- Wallace, A.; Laskowski, R.; Thornton, J. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des, Sel*, 8, pp. 127 –34, 1995.
- Woo, N.; Teng, H.; Siao, T-J., et al. Activation of p75NTR by proBDNF facilitates hippocampal long-term depression. *Nat neurosc.*, v. 8, pp. 1069-77, 2005.

Yamada, K.; Mizuno, M.; Nebeshima, T. Role for brain-derived neurotrophic factor in learning and memory. *Life Sci.*, v.70, n.7, pp. 735-44, 2002.

Yang, J.; Sião, C-J; Nagappan, G., et al. Neuronal release of pro-BDNF. *Nat Neurosci*, v.12, n.2, pp. 113-15, 2009.