Positive Evolutionary Selection of an HD Motif on Alzheimer Precursor Protein Orthologues Suggests a Functional Role

István Miklós[1], Zoltán Zádori[2]*

1 Rényi Institute, Budapest, Hungary, 2 Veterinary Medical Research Institute, Hungarian Academy of Sciences, Budapest, Hungary

Abstract

HD amino acid duplex has been found in the active center of many different enzymes. The dyad plays remarkably different roles in their catalytic processes that usually involve metal coordination. An HD motif is positioned directly on the amyloid beta fragment (Aβ) and on the carboxy-terminal region of the extracellular domain (CAED) of the human amyloid precursor protein (APP) and a taxonomically well defined group of APP orthologues (APPOs). In human Aβ HD is part of a presumed, RGD-like integrin-binding motif RHD; however, neither RHD nor RXD demonstrates reasonable conservation in APPOs. The sequences of CAEDs and the position of the HD are not particularly conserved either, yet we show with a novel statistical method using evolutionary modeling that the presence of HD on CAEDs cannot be the result of neutral evolutionary forces (p<0.0001). The motif is positively selected along the evolutionary process in the majority of APPOs, despite the fact that HD motif is underrepresented in the proteomes of all species of the animal kingdom. Position migration can be explained by high probability occurrence of multiple copies of HD on intermediate sequences, from which only one is kept by selective evolutionary forces, in a similar way as in the case of the "transcription binding site turnover." CAED of all APP orthologues and homologues are predicted to bind metal ions including Amyloid-like protein 1 (APLP1) and Amyloid-like protein 2 (APLP2). Our results suggest that HDs on the CAEDs are most probably key components of metal-binding domains, which facilitate and/or regulate inter- or intra-molecular interactions in a metal ion-dependent or metal ion concentration-dependent manner. The involvement of naturally occurring mutations of HD (Tottori (D7N) and English (H6R) mutations) in early onset Alzheimer's disease gives additional support to our finding that HD has an evolutionary preserved function on APPOs.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: zadori@vmri.hu

Introduction

Human Alzheimer precursor protein (APP) gene was brought to the forefront of scientific interest in the late 80's when protein sequencing of the major component of the amyloid plaques, the amyloid β peptide (Aβ) implicated APP in the development of Alzheimer's disease (AD) [1]. Subsequent genetic studies revealed that mutations in or multiplications of the APP gene alone can cause early-onset AD with cerebral amyloid angiopathy [2]. APP has two homologues, *APLP1* and *APLP2* in vertebrates and orthologues, like Appl (*Drosophila*), and apl-1 (*Caenorhabditis*), all over the animal kingdom. All orthologues and homologues show high sequence and domain homology but the Aβ domain remains a unique feature of the vertebrate APPs [3,4].

APP is a Type-I transmembrane protein with a complex domain organization. So far eight domains were identified on the mammalian APPs, the growth factor like domain, the copper-binding domain, Kunitz-type protease inhibitor domain, the OX2 domain, the glycosylated E2 domain, the unstructured carboxy terminal region of the APP extracellular portion, the transmembrane domain and the short cytoplasmic tail that is involved in transcriptional signaling [5]. Despite intense industrial and academic interest, the physiological and developmental role of APP and the contribution of the different domains to the APP's function are not completely understood.

The human APP gene is ubiquitously expressed not only in glial and neuronal cells but also in almost all tissues that have been examined. The pre-mRNA contains 19 exons and it is alternatively spliced to produce several isoforms. In the brain APP695 is the major component, which compared to the longer versions, is missing the KPI and the OX2 domains on its extracellular portion [6].

APP is localized to many membranous compartments within the cells. It travels through the endoplasmic reticulum and the Golgi apparatus to reach the cell membrane where it is re-internalized in the lysosomes. During this journey the majority of APP is processed on the cell surface by α-secretases which results in the membrane-bound C83 fragment and the soluble APPsa fragment. C83 is cleaved further in its transmembrane region by gamma secretases and leads to the release of the P3 fragments (Aβ$_{17-40/42}$) and the APP intracellular domain (AICD) into the intercellular space and the cytosol, respectively. However, the minority of the APP might be processed on the amyloid pathway by the β secretase complex which generates APPsb and a 16 amino-acid

Author Summary

HD amino acid duplex can be found in the active center of different metallo-enzymes. An HD motif is positioned directly on the amyloid beta (Aβ) fragment and on the carboxy-terminal region of the extracellular domain of the human amyloid precursor protein (APP) and a taxonomically well defined group of APP orthologues (APPOs). The conservation of the HD dyad is not position specific and it cannot be seen in a multiple alignment. Yet we show with a novel statistical method using evolutionary modeling that HD motif is positively selected by evolution on APPOs, despite the fact that HD dyad is underrepresented in the proteomes of all species of the animal kingdom. CAED of all APP orthologues and homologues are predicted to bind metal ions including Amyloid-like protein 1 (APLP1) and Amyloid-like protein 2 (APLP2). Our results suggest that HDs on the APPOs are most probably key components of metal-binding domains, which facilitate and/or regulate inter- or intra-molecular interactions in a metal ion-dependent or metal ion concentration-dependent manner. The involvement of naturally occurring mutations of HD (Tottori (D7N) and English (H6R)) in early onset Alzheimer's disease gives additional support to our finding that HD has an evolutionary preserved function on APPOs.

longer version of C83, the so-called C99 fragment and to a lesser extent it can also cleave within the Aβ domain between Tyr10 and Glu11. The cleavage of the β secretase-generated fragments by γ secretase leads to the release of the AICD into the intracellular compartment and to the generation of $A\beta_{1-40}$, the more neurotoxic $A\beta_{1-42}$, and $A\beta_{11-40/42}$ (see Figure 1) [7,8,2].

Several hypotheses sprung up to explain Aβ's contribution to the etiology of AD, including the amyloid cascade [9], the oxidative stress [10], and the signal transduction hypothesis [11,12]. A Recent model couples Aβ with the loss of ionic homeostasis and the hyperphosphorylation of the Tau protein [13]. Novel experimental data provided additional support to this hypothesis [14]. It seems that all amyloid versions including the P3 fragments are able to oligomerize and the insertion of these oligomers into cellular membranes, by ion channel formation, leads to the perturbation of ionic homeostasis and consequently neurite degeneration and cell death [15,16,17,18].

An HD amino acid duplex has been found in the active center of many different enzymes. Most of these are metalloenzymes like phospholipase A2 [19], metal-dependent phosphohydrolases [20] and metalloproteinases [21]. However, the HD dyad is not restricted to various hydrolyses but it is also an intrinsic part of the active center of the bacterial cycJ/ccmE heme chaperones which are key players in the bacterial cytochrome C biogenesis [22]. In these divergent groups of enzymes HD plays remarkably different roles in the catalytic processes. In sPLA2, H48 (H of HD) functions as a general base and it is assisted by a distal aspartic acid D99 to deprotonate a catalytic water molecule that hydrolyzes the phospholipid ester. The adjacent D49 (D of HD), via its β-carboxyl group coordinates the catalytic Ca2+ cofactor involved in the stabilization of the transition state [23,24,25]. In metal-dependent phosphohydrolases HD coordinates the catalytic metal ion by their side chains as intrinsic part of a metal-binding motif H…HD…D. Mutations of HD eliminate or greatly reduce the enzyme activity as it was shown in the case of the YfbR 5-deoxyribonucleotidase [26] and the tRNA nucleotidyltransferase [27] of E. coli.

In α-secretases, which are members of the adamalysin/ADAM metalloproteinase family, the fully conserved Asp-416 is involved in intramolecular hydrogen bond interactions and directly follows the last hystidine of the zinc-binding consensus motif HEXXHXXGXXH [28].

In cycJ/ccmE proteins H of the conserved HD covalently binds and releases the hem prosthetic group [29]. The mechanism of binding and releasing of hem is not clear but it depends on the interaction of cycJ/ccmE with other protein components of the cytochrome C maturation pathway.

We have identified an HD dyad on the Aβ domain of the mammalian APP proteins. Although position specific conservation is not observed, we show with a novel statistical method using evolutionary modeling that the motif is positively selected along the evolutionary process in the majority of the APP orthologues (APPOs) despite the fact that no other sequence conservation can be recognized on the carboxy-terminal region of the APPOs extracellular domains (CAEDs). In addition, we also show that HD dyads in the proteome of various organisms are under-represented, which further supports the hypothesis that the prevalence of HD in CAEDs is the result of evolutionary selection rather than arbitrary events. The conservation of HD in CAEDs strongly suggests a functional role of this motif, which most likely involves metal coordination or chelation.

Materials and Methods

Proving positive selection of HD motifs

APP orthologues have been collected from the NCBI protein databank with the Blast-P program. The following proteins have been found: Acyrthosiphon pisum, XP_001947569.1; Culex quinquefasciatus, XP_001864483.1; Brugia malayi, XP_001899252.1; Caenorhabditis briggsae, XP_002644641.1; Loligo pealei, ABI84193.2; Aplysia californica, AAT07668.3; Aedes aegypti, EAT42567.1; Drosophila simulans, EDX16764.1; Drosophila yakuba, EDX00795.1; Anopheles gambiae str. PEST, XP_312126.4; Nematostella vectensis, EDO45291.1; Drosophila willistoni, XP_002067462.1; Drosophila persimilis, XP_002027785.1; Drosophila pseudoobscura pseudoobscura, XP_001354498.2; Drosophila virilis, XP_002055698.1; Drosophila grimshawi, XP_001992447.1; Drosophila erecta, XP_001982404.1; Drosophila ananassae, XP_001966309.1; Nasonia vitripennis, XP_001601635.1; Culex quinquefasciatus, XP_001864483.1; Pediculus humanus corporis, XP_002426948.1; Manduca sexta, AAY25024.2; Rattus norvegicus, NP_062161.1; Mus musculus, Q53ZT3; Monodelphis domestica, XP_001373948.1; Equus caballus, XP_001499900.2; Sus scrofa, ABB82034.1; Gallus gallus, AAG00594.1 Canis lupus familiaris, AAX81908.1; Macaca fascicularis, BAD51938.1; Ailuropoda melanoleuca, XP_002920108.1; Oryctolagus cuniculus, XP_002716819.1; Pan troglodytes, AAV74286.1; Callithrix jacchus, XP_002761374.1; Stenella coeruleoalba, AAX81912.1; Xenopus (Silurana) tropicalis, AAH75266.1; Xenopus laevis, AAH70668.1; Pongo abelii, NP_001127014.1; Cricetulus griseus, AAB86608.1; Chelydra serpentina serpentina, AAN04908.1; Apis mellifera, XP_624124.3; Ixodes scapularis, XP_002400744.1; Schistosoma mansoni, CAZ32701.1; Hydra magnipapillata, XP_002154415.1; Neohelice granulata, ACO59955.1; Paracentrotus lividus, CN53783.1; Strongylocentrotus purpuratus, XP_790315.2; Saccoglossus kowalevskii, P_002741027.1; Branchiostoma floridae, XP_002613121.1; Narke japonica, BAA24230.1; Takifugu rubripes, O93279.1; Tetraodon fluviatilis, O73683.1; Danio rerio, NP_690842.1; Tetraodon nigroviridis, CAG05838.1.

```
              1......10.......20.......30.......40.......50.......60.......70.......80.......90.....
Nematostella       PTEKMTEAPTTPSAVHGEEDKNKPHRKHEERERHHGRHGNSKKHRKHDEEDEPENERDFFGEPRSPATFAAVIGLSCGALVIMIVIIVAMAM-R--   } Anthozoa
Caenorhabditis e.  DDDEDSSSSTSSSESDEDEDKNIKELRVDIEPIIDEPA---------SFYRHDKLIQ-SPEVERSASSVFQPYVLASAMFITAICIIAFAITNAR--
Caenorhabditis b.  DDEEDSSSSSSESDEDEDKNIKELRVDIEPIIDEPA---------SFYRHDKLIQ-DSDDESSASTIFQPYVLASAMFITAICIIAFAITNAR--   } Nematoda
Brugia             EDDEDDIEDTDEEPLSSSKSVERELHVEIEPIVSEPVRVHDLPPKPSYARHEPLQHFKMSAEDASASMLKTSTNLLLIFVSAATVIFIVFATVLVR
Nasonia            QQQVQQVHEQQEHHDQQQPSAAASTAVKHSEGLIRAAHSMTHDVSHSEPGYSVRREV-----YHRESRSIYFTMAFAGIALMAATIVGVAVFK--
Apis               QEESDSINDNKDIQQQPEQPAAM------HSEGLVRAAHSLTHDISHSEPGYSVRREM-----YHRESRSVYFTLAFAGIALMAAAVVGVAVFK--
Pediculus          ETMEDEKIVEKDETTVTTAKTLDVLSAVHHN--MPKVAHAQSHDISHSEPSYSLRREI-----YHHESRSIYFTLAFAGIALMAAVIVGVAVFR--
Culex              SPQSTALPTVDDEAVQRAVEEVAAAVAHQE--AEPKMQHVMAHDIGHGEPSYSVRREI-----YGPNGRENRNVYFTLMFAFAVALMGAVFVGVAVAK--
Aedes              SPQSTALPTVDDEAVQRAVEEVAAAVAHQE--AEPKMQHILSHDTGHSEPTFSVRREI-----YGPNGRENRNVYFTLMFAFAVALMGAVFVGVAVAK--
Anopheles          SPQPTALPTVDDEAVQRAVEEVAAAVAHQE--AEPKMQHVLAHDIGHGEPSYSVRREV-----YSSSGRDSKNVYFTVGFAGIALMAAVFVGVAVAK--
Acyrthosiphon      QQLQAQQQPTAIKPDQNAAAAPAAAVQHHDHDSEPKISHAQVHDISHGEPTFFTRNXV------HRDSRNVHLTLAFAGLSLIAAICVGVAVIR--
Drosophila sim.    TKVQTVLPTVDDDAVQRAVEDVAAAVAHQE--AEPQVQHFMTHDLGHRESSFSLRREFAQHAHAAKEGRNVYFTLSFAGIALMAAVFVGVAVAK--
Drosophila yak.    TKVQTVLPTVDDDAVQRAVEDVAAAVAHQE--AEPQVQHFMTHDLGHRESSFSLRREFAQHAHAAKEGRNVYFTLSFAGIALMAAVFVGVAVAK--
Drosophila wil.    TKVQTALPTVDDDAVQRAVEDVAAAVAHQE--AEPQVQHFMTHDLGHRESSFSLRREFA-HSHATKEGRNVYFTLSFAGIALMAAVFVGVAVAK--
Drosophila per.    TKVQTVLPTVDDDAVQRAVEDVAAAVAHQE--AEPQVQHFMTHDLGHRESSFSLRREFA-HAHAAKEGRNVYFTLSFAGIALMAAVFVGVAVAK--
Drosophila pse.    TKVQTVLPTVDDDAVQRAVEDVAAAVAHQE--AEPQVQHFMTHDLGHRESSFSLRREFA-HAHAAKEGRNVYFTLSFAGIALMAAVFVGVAVAK--
Drosophila vir.    TKVQTVLPTVDDDAVQRAVEDVAAAVAHQE--AEPQVQHFMTHDLGHRESSFSLRREFT-HAHAAKEGRNVYFTLSFAGIALMAAVFVGVAVAK--
Drosophila gri.    TKVQTVLPTVDDDAVQRAVEDVAAAVAHQE--AEPQVQHFMTHDLGHRESSFSLRREFA-HAHAAKEGRNVYFTLSFAGIALMAAVFVGVAVAK--
Drosophila ere.    TKLQTVLPTVDDDAVQRAVEDVAAAVAHQE--AEPQVQHFMTHDLGHRESSFSLRREFAQHAHAAKEGRNVYFTLSFAGIALMAAVFVGVAVAK--
Drosophila ana.    TKVQTVLPTVDDDAVQRAVEDVAAAVAHQE--AEPQVQHFMTHDLGHRESSFSLRREFSHNAHAPKEGRNVYFTLSFAGIALMAAVFVGVAVAK--
Loligo             TEARPEQQEDIKTPGFDSETFEDERPAIAEQTIHDLPNHRKKGYIAHVQQQPMIADEVSLDNLYANSHANSVLGIAIGGVVVFIIIVVAVVMLK--   } Mollusca
Aplysia            SPDDDIRIDDAGDFDLSSEQVDEEGDDVSEDE-HDY--ERGSQFIAHRMDDKMHVRQGFAESAATSSQMGSTIGIALGGVSVFVIIVVAVFMVK--
Xenopus tro.       ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDSEYRHDAAYEVHHQKLVFFAEEVGSNKGAIIGLMVGGVVIATVIVITLVMLK--   } Amphibia
Xenopus lae.       ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDSEYRHDTAYEVHHQKLVFFAEEVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Chelydra           -------------------------LTNVKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--   } Reptilia
Gallus             ANTENEVEPVDARPAADRGLTTRPGSGLTNVKTEEVSEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--   } Aves
Monodelphis        ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDMNSNKGAIIGLMVGGIVIATVIVFTLVMLK--
Oryctolagus        ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Mus                ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFGHDSGFEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Cricetulus         -------------------TPGSGLTNIKTEEISEVKMDAEFGHDSGFEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Rattus             ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFGHDSGFEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Stenella           -----------------------GSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDMGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Ailuropoda         ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Canis              ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--   } Mammalia
Sus                ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Equus              ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Callithrix         ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEHVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Macaca             ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATV----------
Pongo              ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Pan                ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--
Homo               ANTENEVEPVDARPAADRGLTTRPGSGLTNIKTEEISEVKMDAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--

                                        β    β'   α                              γ

Narke              ANTENEVEPVDARPAPDRGLTTRPGSGLTGIKTEEIAELKMETEFQQDSGYEVHHQKLVFFPKDVGSNKGAIIGLMVGGVVIATVIVITLVMLK--   } Chondrichthyes
Takifugu           ANTENHVEPVDARPIPDRGLPTRP---VSSLKLEEMPEVRTETDKRQSAGYEVHHQKLVFFADDVGSNKGAIIGLMVGGVVIATVIVITLVMLR--   } Actinopterygii
Tetraodon          LNTENQVEPVDSRPTFERGVPTRP---VTGKSMEAVPELRMETEDRQSTEYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIATVIVITLVMLR--
Danio              ANTDNHVEPVDARPIPERGLPTRP---------EIPKVRLDIEERHNAGYDVRDKRLMFLAEDMGSNKGAIIGLMVGGVVIATVIVITLVMLR--
Branchiostoma      VTPQHLQKFAASKADRPPQAASIVSQQMMNHVQWSCYSSSNETRQCSCYQREHYLCLRVFHMSHIT-NLTTLS-LTVAAAIIAVIIIVALIVLR--   } Cephalochordata
Saccoglossus       EVMEMLMTTIAPDLTTYDEETMEDESILPVQQKLMASKFDASKDSKFKVQESFNKPFVVKKKMTYPHKSFLALGLVGGAVAVATVVIIALVIVR--   } Hemichordata
Neohelice          ADFGLRAASSAPNASTPNASARPEPLLVHHAED-VQVAHAQHHQISHSQASYSVTRES-----SGSESRGVYFTLAFAGVALMAMMVVGVAVVLK--   } Crustacea
Ixodes             QPSGDSPSSPSSSSSSSSSSSPLPQQGAQLLVHPTPRVAHAQGQLLHHSEGSYNTMRKE--GLQKFKWNGSVYITLAFAGVALLTALLVGAVLLR--   } Chelicerata
Paracentrotus      YLDEELTAPIETEDVKPQTLIIAQKPKAKVEKSAAAARMEKQRLVNRQDVNSIDNS-VTKKRMQLRSEPMMAFGLACGVLAVATILVIAVLIVR--   } Echinodermata
Strongylocentrotus YLDEELTAPLDSDVVKPQTLIIAQKPKAKVEKSAAAARMEKQRLVGRQDVNSIDNS-VTKKRMSLRYEPMMAFGLACGVLAVATVLVIAVLIVR--
Schistosoma        STVDYSSSLSLNQNQRTPHINSSEMIENKTTVQFGSYQNISVVMLHSDKVLNTSNSSTVNLSTQIHHFKTTYI-LLILLGIVLCFICLFIILRR--   } Platyhelminthes
Hydra              DDEELVAVKKLKVQTTVDTNIVDPSTTTVFPQNLALELKSSDEDVANVVQEEPKNNEHH--RSKRSVLLTAILGLSCGIVVVMLFVVVALIM-R--   } Hydrozoa
```
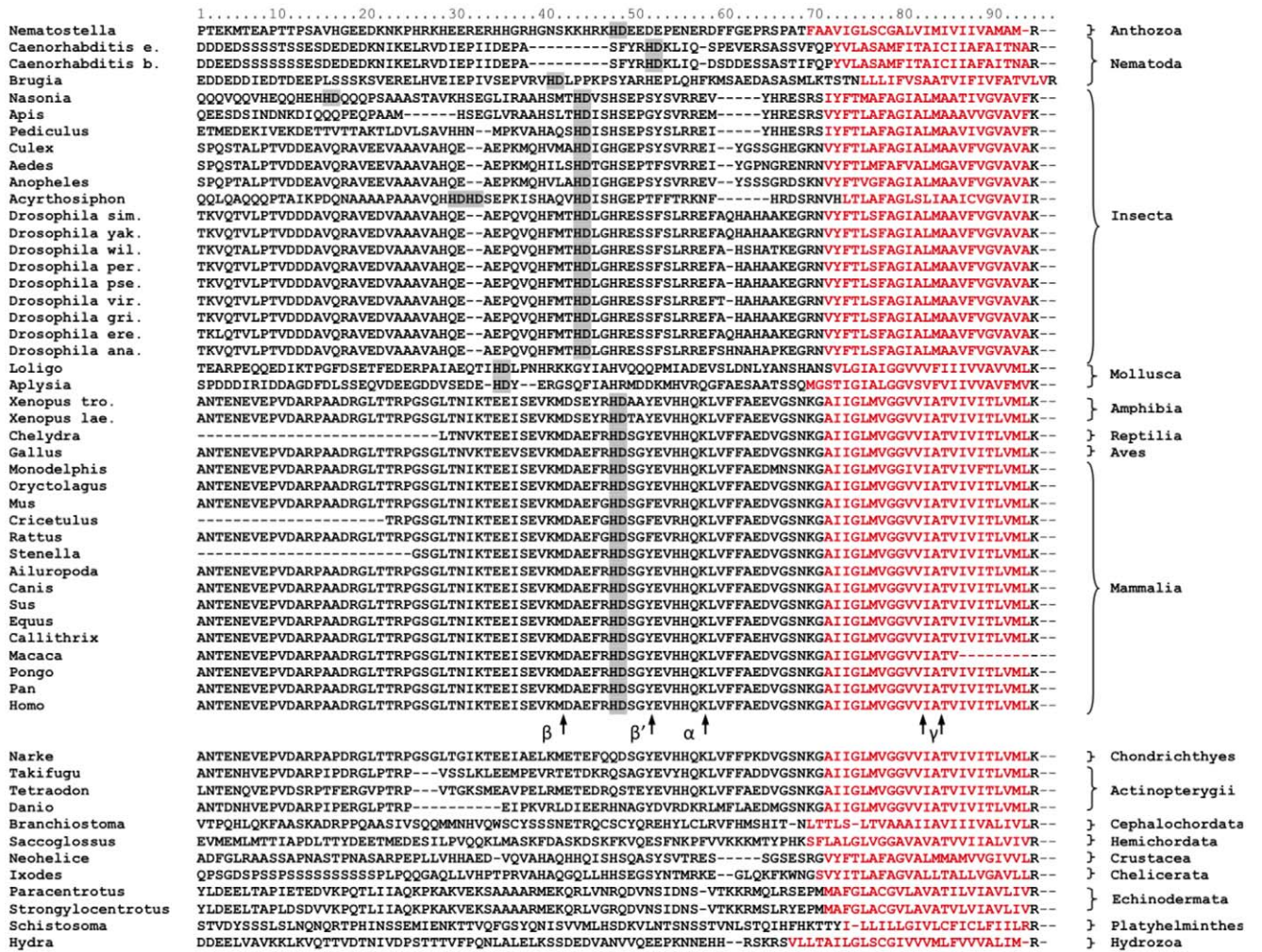
Figure 1. Multiple alignment of the membrane-proximal regions of CAEDs and the transmembrane helices of the APPOs. The predicted transmembrane domains are in red. From the CAEDs only the regions homologous to the predicted metal-binding site of the human APP are shown. The HD dyads are highlighted by gray. Digestion sites of the human α-, β-, and γ-secretases are marked by arrows.
doi:10.1371/journal.pcbi.1002356.g001

The retrieved proteins were globally aligned with MultAlin [30]. The transmembrane helices of the proteins have been identified by HMMTOP [31] and regions comprising the last 70 amino acids of the CAED sequences and the membrane spanning helices were realigned again with MultAlin. Different regions of this alignment (Figure 1) were used as the inputs for MrBayes 3 [32], a program that samples evolutionary trees from a Bayesian distribution with a Markov chain Monte Carlo method. We used the default parameters of MrBayes, and the convergence of the chain was checked by the log-likelihood trace. We discarded the first half of the chain as burn-in and the consensus tree of the sampled trees was obtained by the consensus network method [33] using SplitsTree 4.0 [34]. There were no ambiguities in the topologies of the trees. As an example, a tree is shown in Figure 2, created from the alignment corresponding to the last 70 amino acids of human CAED.

We implemented a program in the Java 1.6 language that takes an evolutionary tree and a sequence labeling its root, and evolves the sequences on the tree according to a substitution model represented with a continuous time Markov model. The Markov model is given by its rate matrix, Q, we used the rate matrix being equivalent with the BLOSUM62 matrix, see [35] for details. The exponent of the matrix contains the so-called transition probabilities. For example, the entry in row k and column l contains the probability that a site is in a particular amino acid a_k after evolutionary time t, given that the site was in amino acid a_l in the beginning. The exponet of the matrix by definition is

$$e^{Qt} = \sum_{i=0}^{\infty} \frac{(Qt)^i}{i!} \qquad (1)$$

but for technical reasons, we use the diagonalized form of the matrix. If

$$Q = W\Lambda W^{-1} \qquad (2)$$

where Λ is a diagonal matrix containing the eigenvalues of Q, then

$$e^{Qt} = We^{\Lambda t}W^{-1} \qquad (3)$$

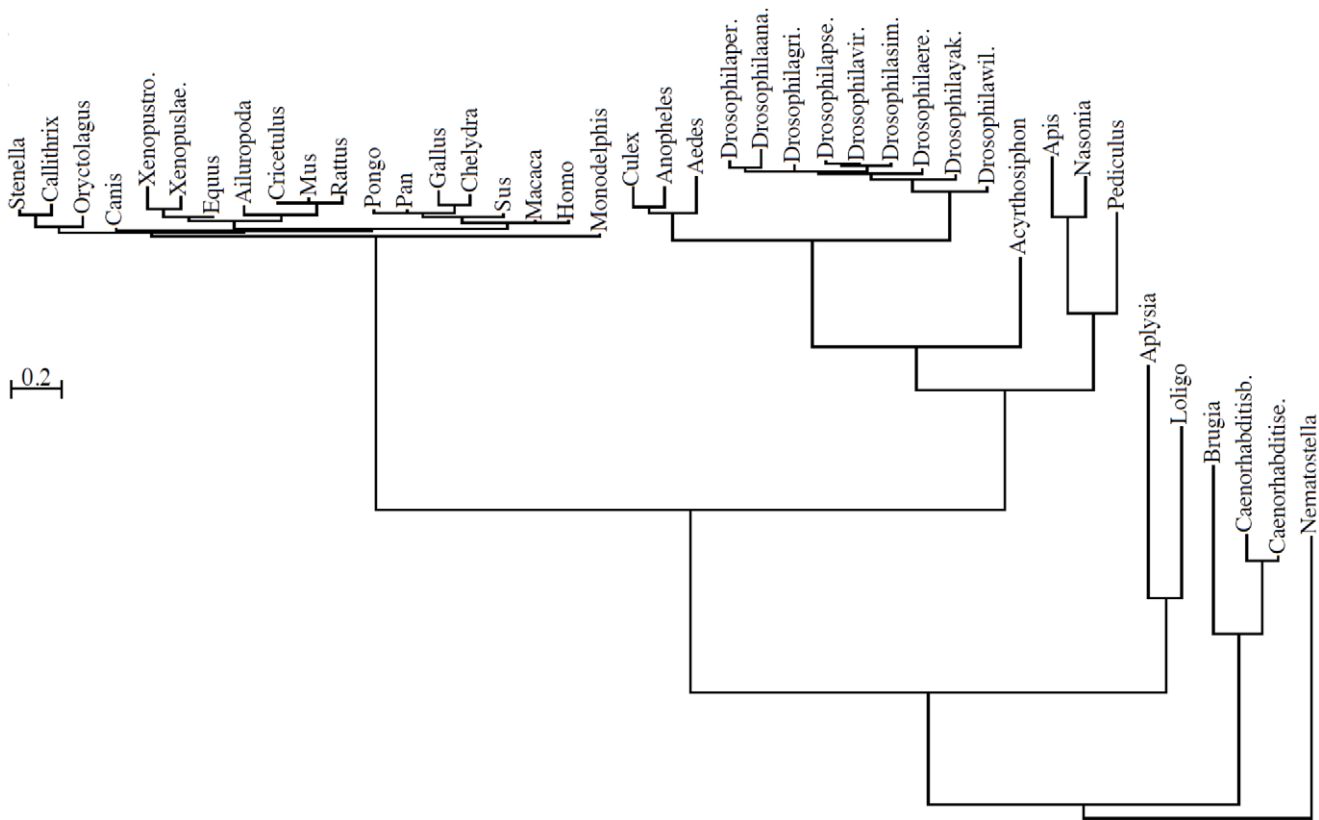which is much easier to calculate since

Figure 2. Consensus tree of the membrane-proximal regions of CAEDs containing HD motifs. The tree was calculated from region 1–70 of the alignment shown in Figure 1 by SplitsTree 4.0 using the consensus network method. Bar represents 0,2 PAM distance.
doi:10.1371/journal.pcbi.1002356.g002

$$e^{\Lambda t} = \begin{pmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & e^{\lambda_{20} t} \end{pmatrix} \quad (4)$$

More details about the mathematical background of the calculation method are given elsewhere [36–38].

Each site in the sequence is evolved independently of the other sites, and the evolution on each branch of the evolutionary tree is also independent of the evolution on other branches. We took the evolutionary tree generated by SplitsTree 4.0, put the *Nematostella vectensis* CAED to the root, and let the sequences evolve according to a substitution model being equivalent to the BLOSUM 62 score matrix. Namely, for each site, and each edge of the tree from the root towards the leaves, we took the amino acid at the incoming vertex of the edge. We calculated the exponent of the rate matrix using the evolutionary time assigned to the edge in question, then generated a random amino acid for the outgoing vertex of the edge. If the amino acid at the incoming vertex was a_l, than the amino acid at the outgoing vertex of the edge was generated from the distribution taken from the lth column of the exponentiated rate matrix. We ran the program simulating the evolution on the tree 10 000 times, and 10 000 sequence sets containing the sequences generated at the leaves of the tree were collected.

The number of every amino acid dyad in the generated sequences was counted, and the distribution of the sequence sets with different number of sequences containing any given dyads was calculated. The observed number of CAEDs with HD dyads (41 altogether) was compared with the empirical distribution of the 10 000 sequence sets calculated to the HD dyad to test the following hypothesis:

H$_0$. The amino acids in the CAEDs evolved neutrally; there is no selection force to maintain at least one HD motif in the domain.

H$_1$. The amino acids in the CAEDs did not evolve neutrally; there is a selection force to maintain at least one HD motif in the domain.

The p-value is the probability of a value at least as high as the observed value (41 HD containing "real" sequences) assuming the H_0 hypothesis. The neutral evolution hypothesis can be tested for any dyad, and thus, similar p-values were calculated for other dyads, as well.

Inferring the HD diamino-acid distribution in proteomes

We downloaded the Uniprot/Swisprot database, 2011-05-31 release (ftp://ftp.ebi.ac.uk/pub/databases/uniprot/knowledge-base/) in fasta file format. Derivative databases were generated for all species with more than a thousand deposited sequences. For each database, the empirical distribution of single amino acids as well as the empirical distribution of diamino-acids was calculated. From this, the log-odds were calculated for each pair of amino acids. By definition, the log-odds for amino acid pairs a and b is:

$$s(a,b) := \log\left(\frac{\pi(a,b)}{p(a)p(b)}\right) \quad (5)$$

where p(a) and p(b) are respectively the probabilities of amino acid a and b in the single amino acid distribution, and $\pi(a,b)$ is the probability of the ab diamino-acid motif in the diamino-acid distribution. (See Table S1).

Prediction programs

For secondary structure prediction the PSIPRED [39] and Jpred [40] software were applied. PSIPRED uses position specific scoring matrices generated by PSI-BLAST to predict protein secondary structure by a two-stage neural network. The program proved to be the most accurate (76.5% to 78.3%) among all investigated methods in the third Critical Assessment of Techniques for Protein Structure Prediction experiment (CASP3) [39].

Jpred runs also a neural network predictor Jnet v3.0, which combines PSI-BLAST position scoring matrix with hidden Markov profiles and achieved a secondary structure prediction score of 81.5% in blind experiments. In a validation test the two programs produced largely overlapping results, however on a portion of the data set one or the other programs gave more accurate prediction [40], so combined application of the two programs and comparison of their results might sometimes be beneficial.

Metal-binding was predicted by the SVMProt server [41] and the MetalloPred program [42].

MetalloPred classifies proteins from sequence derived features (like amino acid composition physicochemical properties and pseudo-amino acid composition) by using a three level cascade of neural networks. The 1st layer of the cascade is for finding metalloproteins, the 2nd layer for the main functional classes (e.g transition metal); and the 3rd layer for identification of the bound metal (e.g. zinc). The accuracy of the program at the first level is reported to be >80%, while the overall accuracy for the correct metal recognition is higher than 60%.

The SVMProt server runs support vector machine prediction systems to predict metal-binding proteins with 10 metal-binding classes (e.g. sodium-binding and zinc-binding, etc). It recognized metal-binding domains and multi-domain metal-binding proteins with more than 80% accuracy in validation tests.

Results

Alignment and sequence similarity

Screening of protein databases revealed an HD dyad on the mammalian APP proteins. The motif is positioned directly on the amyloid beta fragment of human APP (amyloid beta conventional numbering H6, D7; APP conventional numbering H677, D678) and it seems to be conserved among not only mammals but four legged vertebrates (tetrapoda), too. No APLP1 or APLP2 orthologues of any animals contain the motif in their functionally homologous region (data not shown).

To clarify the functional and evolutionary significance of the HD motif in APP proteins, first the available APPOs were collected from protein databanks and their CAEDs and the transmembrane (TD) domains were aligned (Figure 1). The alignment revealed that the presence of Aβ domain, embedded in the CAED and TD, is indeed restricted to vertebrate APPs as it was earlier recognized by others [3,43]. The uniqueness of Aβ in vertebrate APPs is the direct consequence of the fact that the CAEDs and TDs show little sequence conservation between taxonomically divided larger groups like insects, vertebrates and nematodes. In-group conservation of CAED inversely correlates with diversity, e.g. the CAED of insects are less conserved than the CAED of vertebrates. So, in general it can be stated that the more related the animals, the more similar their CAED and TD are on

the APPOs. However, there are some exceptions. Interestingly, the CAED and Aβ of the cartilaginous fish *Narke Japonica* [44] show significantly higher homology to human APP than to that of the more related bony fishes [45].

Motifs and secondary structure predictions

Despite the lack of conservation in the animal kingdom, the majority of the CAEDs contain an HD dyad in their last 70 amino acid region (CAED$_{C70}$). In spite of intensive search we were unable to find any local conservation around the dyad or a conserved amino acid pattern in the CAEDs, which would include the HD motif. In human Aβ HD is part of a presumed RGD-like integrin-binding motif RHDS [46,47] which was shown to promote cell adhesion and α5βl integrin binding of the soluble form Aβ but not of APP or the fibrillar form of Aβ [48]. The selectivity of integrin binding toward soluble Aβ could be explained with the structural requirements of RGD binding. RGD must be on a tongue-like loop to fit into the 'well-shaped' ligand-binding pocket of the integrin where the carboxylate of the aspartic acid coordinating a zinc ion and the basic arginine moiety through a salt bridge anchors the ligand [49,50,51]. It seems plausible that RHD of the soluble Aβ can fold into loop while RHD of the more structured APP and the fibrillar form cannot. It also follows from the mechanism that HD, with high probability, cannot be evolutionary preserved on CAEDs to promote integrin binding (or at least not in an RGD-like manner) because the arginine residue in the RHD motif of the human Aβ does not demonstrate reasonable conservation, and it is not conserved even among mammals (GHD in rodents). In fact, the RXD motif is missing from the majority of non-vertebrate CAEDs.

Dissimilar sequences frequently share similar secondary structures and folds [52], so we have investigated with secondary structure prediction programs whether CAEDs and especially the HD surrounding areas share any common secondary structures. Aβ has a propensity for conformational change; its actual structure largely depends on the interacting environment. Therefore it can be considered as a so called "dual personality fragment" [53] (an extended concept of "chameleon sequence" [54,55]). The monomeric Aβ is unordered in aqueous solution and takes up a helical structure in membrane-mimicking media [56–61]. By contrast, in amyloid fibrils and ion channel-forming aggregates residues 18–42 adopt a β-strand–turn–β-strand motif [62,18]. Residues 16–23 compose a discordant helix which seems to be critical for fibril formation [63]. Contrary to the Aβ, not much is known about the CAEDs structure in APPOs except that human CAED was predicted to be unordered [5]. Our prediction results have been inconclusive. Human CAED$_{C70}$ is predicted to be largely unordered coil, in which the HD is bracketed by a short helix and a β strand. This prediction does not exactly match any Aβ experimental data, which is not surprising, considering that predicting secondary structure of dual personality fragments is intrinsically difficult. The divergent CAEDs did not show uniformity in secondary structural features either on the full length CAEDs or in the HD surrounding areas, though in the CAEDs of vertebrates and some insects HD is located at the carboxy-terminal of a (5–10 amino acid length) helix (Figure S1).

Taxonomic segregation of CAEDs with HD motif

The presence of HD on different APPOs is far from being arbitrary and shows a progressive taxonomic distribution in certain animal groups. Though it appears first in the Anthozoa class of the Cnidaria phylum, it is absent from the APPOs of the Hydrozoa class, the Platyhelminthes and Echinodermata phyla. Besides tetrapods, it can be found in all the available APPOs of

insects, mollusks and nematodes, and it is completely missing from the members of the related primordial taxons of the deuterostomia and arthropoda lineages like cephalochordates, bony and cartilaginous fishes, crustaceans and chelicerate arthropods (Figure 1). The evolution of tetrapods from sarcopterygian fish could be dated in the Late Devonian period [64] while insects were separated from other arthropods around 400 million years ago in the early Devonian, more than 150 million years later than crustaceans and chelicerate arthropods [65,66]. Based on these data it is tempting to speculate that the HD dyad independently evolved in the naïve ancestors of tetrapods and insects and it is sustained in the CAEDs of the species of these and other taxons by a similar if not identical evolutionary force. However, the division of CAEDs regarding the presence of HD and the lack of recognizable sequential or structural conservation in CAEDs around the HD dyad forced us to investigate the possibility that the appearance of these amino acids in the APPs of different animals would be the result of random or arbitrary evolutionary events.

The HD motif is underrepresented in the proteomes, but overrepresented in the CAED domains

First we examined whether the frequent occurrence of the HD on CAEDs could be the result of overrepresentation of HD in the biota and most importantly in the animal proteomes.

The log-odds of the HD motif as defined in Equation (5) in the vast majority of investigated proteomes resulted in negative values. This indicates that the HD motif is underrepresented in the biota in general, as well as in specific species. Among single-cell organisms HD frequency fluctuates and the log-odds can even take positive values, while in multi-cell organisms, regardless of their taxonomical place, it always remains negative. Based on the presently available data, there seems to be an inverse correlation between the log-odds and the evolutionary development of the phyla in the animal kingdom. In the investigated species the values decrease from primitive Bilateria towards modern Bilateria, and they reach the minimum in vertebrates (Figure 3). It is worthwhile mentioning that in the human proteome the number of HDs lags farthest behind the expected value calculated by the amino acid composition indicating that HD has the smallest log-odds among all amino acid dyads (Table S1).

On the other hand, the log-odds value of the HD motif in the $CAED_{C70}$s is 1.355. Namely the HD motif is overrepresented in the $CAED_{C70}$s, its occurrence is much more frequent than the independent distribution of amino acids would indicate.

Positive selection of HD motifs

We have also developed a computer program to study whether neutral evolutionary forces are able to produce such overrepresentation of HDs on the CAEDs. The program takes an evolutionary tree (e.g. the tree is shown in Figure 2) and a sequence labeling its root, and evolves the sequence at the root on the tree by keeping the original distances among the leaves according to a substitution model (a continuous time Markov model) representing neutral evolution. After a user defined number of iteration of the neutral evolutionary process, the program calculates the number of any amino acid dyads (e.g. number of HD) and the number of sequences containing a given dyad (e.g. 26 HD can be on 22 sequences if some of the sequences contain more than one HD) in every set of the computationally evolved sequences.

As an input tree, first the evolutionary tree of the $CAED_{C70}$s was chosen (generated from region 1–70 of the alignment (see Figure 1)) by taking the *Nematostella* sequence as a root. We have

chosen the sequence of the simplest organism to evolve because this way the direction of the evolution in the simulated processes gives the closest approximation of the reality.

As shown in Figure 4, after 10 000 iterations the distribution of the evolved sequence sets with different number of HD motif containing sequences is not unimodal. This is due to the correlation amongst the sequences labeling the leaves of the evolutionary tree. Indeed, if a motif is represented in one of the sequences, the probability that it is also represented in its closely related homologues is higher. In more than 33% of the cases (9,12%+12,81%+11,34%) from the 10 000 simulations, the HD dyad occurred only in no more than two evolved sequences (Figure 4) and never occurred in all 41 computationally evolved sequences.

Therefore we conclude that we can reject the neutral evolution hypothesis with a p value smaller than 0.0001, and HD is kept in the CAEDs by a selective evolutionary force. The cross-validation showed that three dyads (KM, EP and HQ) amongst the possible 400 ones were significant at $p = 0.01$ level (Table S2 Panel A,). This result is comparable to the expected 4 ones (1% of the 400).

Though a large number of $CAED_{C70}$ were tried as input (among them the *Human* and *Drosophila* sequences) to test the effect of the input sequence on the final outcome, in every case the p value of HD dyad remained smaller than 0.0001. Using different trees, generated from different parts of the alignment, can have an influence on the amino acid composition of the evolved sequence sets and consequently on the distribution of the dyads. So we tested several trees derived from shorter and longer regions of the alignment, yet the P value for HD never reached 0.0001, (Table S2 Panel B, and C,) showing that the result is not the consequence of an arbitrary choice.

The lack of position-specific conservation raises the question of how the HD dyad can be kept by a selective evolutionary force without position-specific conservation. In our computer simulations, more than 20% of the simulated sequence sets contained at least one sequence with 2 or more HD dyads, hence, we conclude that the probability for the emergence of a new HD dyad is relatively high. If the evolutionary force is only for maintaining at least one of the HD dyads, then the deletion of the old HD dyad will not be prevented by selection just as it happened in the majority of the modern sequences.

To test further the hypothesis that HD motifs might appear by random mutations, we repeated the simulation of neutral evolution, but now we put the *Hydra magnipapillata* sequence to the root, which does not contain a HD dyad. The distribution of the number of sequences containing a HD motif still followed a multimodal distribution similar in shape to the one on Figure 4, but the distribution was shifted towards smaller values, with an average number of 3.32 HD containing CAEDs/sequence set in contrast to 9.68 HD containing CAEDs/sequence set when we put the HD containing ancestor to the root (data not shown).

Metal-binding prediction

From a structural point of view, an important recognizable common feature of the $CAED_{C70}$s is that despite the lack of sequence similarity they are rich in metal coordinating amino acids (His, Glu, Gln Asp, Asn, Tyr, Ser, Thr Arg, Lys). Certain combinations of these amino acids like EE ED and DD are also relatively frequent on $CAED_{C70}$s (they occur on 34, 37 and 17 protein segments of the 53 CAED respectively), though their occurrences stay below that of the HD. These dyads are also able to coordinate metal ions. EE was reported as part of manganese and nickel coordinating motifs [67,68]. ED and DD bind calcium [69,70] magnesium [71,72] and manganese [73,74]. Interestingly,
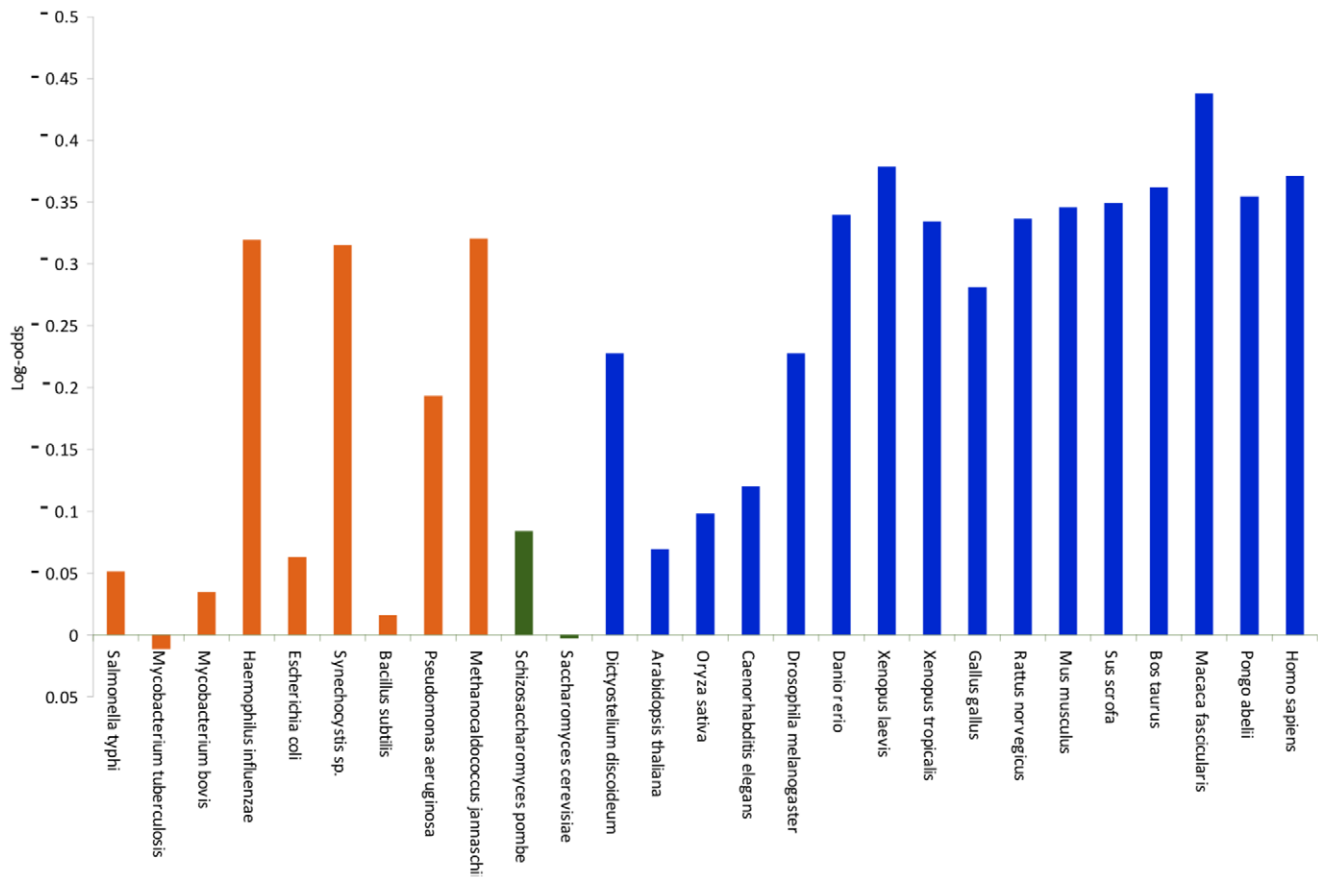
Figure 3. The log-odds values of the HD dyad in the proteomes of several organisms from the biota. Orange, green and blue columns represent prokaryotes, single-cell eukaryotic and multicell eukaryotic organisms respectively.
doi:10.1371/journal.pcbi.1002356.g003

other combination of E, D and H amino acids H-7X-H, E-2X-H, H-3X-H, E-4X-D, which can be found in several metal-binding motifs [75] are also frequent (42, 34, 18, 22 occurrence respectively) on CAEDs. These observations prompted us to investigate *in silico* whether the CAEDs could have metal-binding capabilities. Surprisingly, the last 60–70 amino acid of all CAEDs were predicted to be metal-binding domains, by 2 different prediction programs, regardless of the presence or absence of the HD motif (Table S3). Both programs Metallopred and SVMProt achieved more than 80% accuracy in validation tests [41,42] so the chance that all of these sequentially non-homologous protein segments, with similar functions, would be falsely predicted to have metal-binding capabilities seems to be marginal. Furthermore, the carboxy-terminal region of the extracellular domain of APLP1 and APLP2 orthologues is also predicted to bind metal ions, which suggests that metal-binding capability near the transmembrane anchor are evolutionary maintained and it is indispensable in order that APP orthologues and homologues exert their normal biological functions. This is coincident with the observations that amyloid deposits contain high levels of copper, iron, and zinc [56,76] and natural Aβ is a metalloprotein [77]. What is more, investigations of the metal-binding properties of the Aβ revealed that both amino acids of HD motif (H6 and D7) are involved in metal coordination. The first sixteen amino acids of human Aβ can bind zinc by inter and intramolecular coordination [78,79,80] and H6 contributes to the latter one. Moreover age-related isomerization and racemization of D7 result in zinc dependent oligomerization of Aβ(1–16) peptide and it causes

conformational change in the His6–Ser8 region of Aβ(1–16) which retains its zinc-binding capacity by the involvement of L-iso-D7 [56,81]. Molecular dynamic simulations on Aβ models and EPR studies are also involving unmodified D7 in inter- [82] and intramolecular metal coordination [83] and oligomerization [82].

Taken together, these data indicate that H and D may be involved in metal coordination not only in human Aβ but also on APPOs, and their evolutionary selection can be related to this function.

Discussion

Investigation of the CAEDs of evolutionary distant animals revealed that despite the low sequence similarity, the majority of them contain an HD dyad and their membrane-proximal 60–70 amino acid regions are predicted to bind metals. The HD-containing CAEDs belong to the species of well-defined taxonomic groups such as tetrapodes, insects, mollusks and nematodes. We have shown using an evolution model system that although HD is negatively selected in the proteome of different animals, the presence of the HD dyad on CAEDs is most likely the result of positive selection. We want to emphasize that the conservation of the HD dyad is not position specific; hence its conservation cannot be seen in a multiple alignment. However, the positive evolutionary selection of HD has been proved by statistical testing of the sequences. Under the neutral evolution hypothesis, namely, assuming no selection force for maintaining the HD dyads, the probability of the observed abundance of the HD dyads is less than 0.0001.
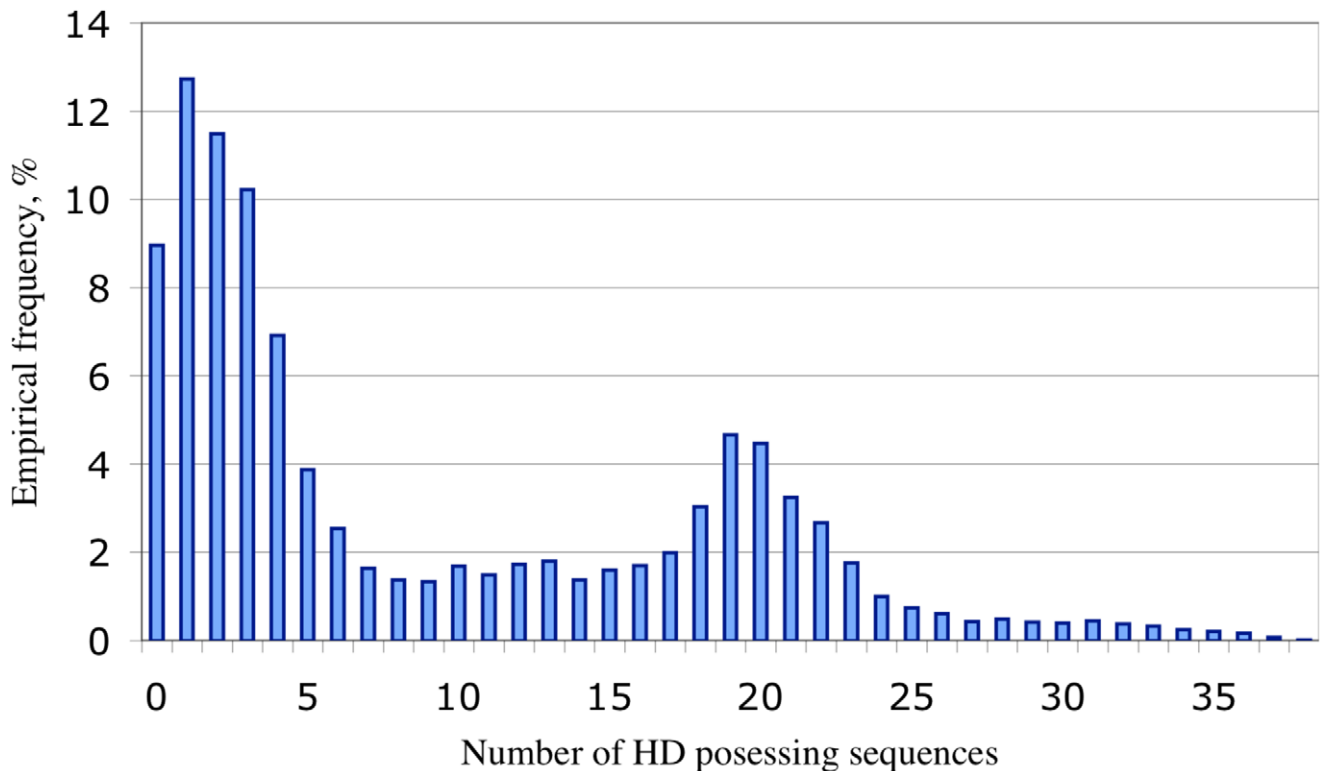
Figure 4. The empirical H_0 distribution of 10 000 sequence sets assuming neutral evolution. Every column represents a group of sequence sets with a certain number (0–41) of HD containing sequences. Numbers on the X axis correspond to the number of HD containing sequences in the group. Groups with 39–41 HD containing sequences have 0 member and they are not indicated on the X axis. Labels on the Y axis show the size of the groups in percentage of the total 10 000 sets.
doi:10.1371/journal.pcbi.1002356.g004

Computer simulations showed that the emergence of an HD dyad in the CAED sequences is likely even under neutral evolution. Two of the CAEDs contain more than one HD dyads. According to the simulation the probability of observing such multiple occurrences in at least one of the modern sequences is over 20% even in the case of neutral evolution. The probability that at least one of the evolving intermediate sequences contained multiple dyads is even higher. Without the selection force, these appearing HD dyads could mutate, thus the sequence could lose at least one, or even all of them. In case of a selection force, at least one of the HD dyads is preserved in the sequence. However, this dyad might be the one that appeared by random mutations, and the older HD dyad might be deleted. This scenario could explain why we see HD dyads in the majority of sequences without site-specific conservation. Similar migration of functional elements have already been described for transcription factor binding sites in DNA promoter regions as the so-called binding site turnover [84,85,86]; however, to the best of our knowledge, this is the first time that such motif turnover is described for proteins.

The evolutionary selection of HD strongly suggests a functional role of the motif in APPOs, which is probably related to metal coordination. As the HD motif is part of the extended catalytic network of several enzymes, the contribution of the CAEDs' HD to the formation of a catalytic center through inter- or intramolecular interaction may not be excluded. However, the lack of conservation in the vicinity of HD, the variable position of the motif on the CAEDs and the multiple copy occurrences in certain proteins indicate a low probability for this supposition. The same arguments which oppose enzymatic activity, together with the fact that membrane-proximal region of the CAEDs seem to

have metal coordination capabilities, rather suggest that HDs on CAEDs could be key components of metal ion-coordinating domains which facilitate and/or regulate inter- or intramolecular interactions in a metal ion-dependent or metal ion concentration-dependent manner. This notion is supported by the findings that Aβ has metal-binding capability [56,76–83], structural plasticity [57–60] metal and metal concentration dependent propensity for structural changes [56,80,82] and metal ions (Cu2+ and Zn2+) facilitate the intermolecular contact between Aβ peptides [82,87] in which H and D are involved by metal coordination [81,82].

APP and its derivatives interact with large number of proteins. Aβ binds its homologous sequence on APP and also facilitates the oligomerization of the β-secretase cleaved APP C-terminal fragment, C99 [88]. Obviously, mammalian CAEDs directly interact with α- and β-secretases (as they are digested by them) and likely with some member(s) of the γ-secretase complex (all components are membrane proteins and have extracellular parts). Besides the well-studied cytoplasmic carboxy-terminal interacting adaptor proteins [8,89,90] large numbers of other extracellularly interacting APP partners were identified with poorly characterized binding features [91]. Both intracellular and extracellular interactions modulate APP processing [92,93]; consequently, their perturbation could lead to elevated production of neurotoxic Aβ species and the development of AD.

If HD is involved in any molecular interaction, which influences APP processing or has any other function that influences the development of AD, then certain mutations of HD may facilitate the manifestation of AD. In fact, there are reports supporting the biological significance of these amino acids in the development of AD. Naturally occurring mutations of HD are involved in the

early onset of familial AD in cases from Japan (D7N) [94] and England (H6R) [95]. These observations provide additional support to the functionality of HD on APPOs.

Though substantial amount of data has already accumulated about the pathogenesis and development of AD, finding the cure may require greater knowledge about the physiological role of APP. We hope that our results can stimulate new investigations and contribute to the better understanding of APP's involvement in the development of AD.

Supporting Information

Figure S1 Secondary structure predictions of some selected CAED$_{C70}$.
(PDF)

Table S1 The log-odds values of the amino acid dyads in the proteomes of several organisms from the Biota. Table A, shows the number of the amino-acids in the proteomes; Table B, shows the log-odds values which are calculated by equation 5. The first amino acids of the dyads are represented on the vertical axis while the second amino acids are represented on the horizontal axis.
(PDF)

Table S2 The p values of the different amino acid dyads in the neutral evolution simulation. The values of Panels A, B, and C were calculated from different regions (1–70, 10–54 and 1–96, respectively) of the alignment shown in Figure 1. Table A, shows the p values of 10 000 simulated runs. Table B, contains the sum of amino acid dyads in the examined regions of the 41 CAEDs. The first amino acids of the dyads are represented on the vertical axis while the second amino acids are represented on the horizontal axis.
(PDF)

Table S3 Metal-binding prediction results on the CAEDs of APPOs.
(PDF)

Author Contributions

Conceived and designed the experiments: ZZ IM. Performed the experiments: ZZ IM. Analyzed the data: ZZ IM. Contributed reagents/materials/analysis tools: ZZ IM. Wrote the paper: ZZ IM.

References

1. Kang J, Lemaire HG, Unterbeck A, Salbaum JM, Masters CL, et al. (1987) The precursor of Alzheimers-disease amyloid-A4 protein resembles a cell-surface receptor. Nature 325: 733–736.
2. Thinakaran G, Koo EH (2008) Amyloid precursor protein trafficking, processing, and function. J Biol Chem 283: 29615–29619.
3. Walsh DM, Minogue AM, Sala Frigerio C, Fadeeva JV, Wasco W, et al. (2007) The APP family of proteins: similarities and differences. Biochem Soc Trans 35: 416–20.
4. Bayer TA, Cappai R, Masters CL, Beyreuther K, Multhaup G (1999) It all sticks together–the APP-related family of proteins and Alzheimer's disease. Mol Psychiatry 4: 524–8.
5. Kong GK, Adams JJ, Harris HH, Boas JF, Curtain CC, et al. (2007) Structural studies of the Alzheimer's amyloid precursor protein copper-binding domain reveal how it binds copper ions. J Mol Biol 367: 148–61.
6. Turner PR, O'Connor K, Tate WP, Abraham WC (2003) Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory. Prog Neurobiol 70: 1–32.
7. Wilquet V, De Strooper B (2004) Amyloid-beta precursor protein processing in neurodegeneration. Curr Opin Neurobiol 14: 582–8.
8. King GD, Turner RS (2004) Adaptor protein interactions: modulators of amyloid precursor protein metabolism and Alzheimer's disease risk? Exp Neurol 185: 208–19.
9. Hardy J (2006) Has the amyloid cascade hypothesis for Alzheimer's disease been proved? Curr Alzheimer Res 3: 71–3.
10. Bennett S, Grant MM, Aldred S (2009) Oxidative stress in vascular dementia and Alzheimer's disease: a common pathology. J Alzheimers Dis 17: 245–57.
11. Pimplikar SW, Nixon RA, Robakis NK, Shen J, Tsai LH (2010) Amyloid-independent mechanisms in Alzheimer's disease pathogenesis. J Neurosci 30: 14946–54.
12. Maccioni RB, Farías G, Morales I, Navarrete L (2010) The revitalized tau hypothesis on Alzheimer's disease. Arch Med Res 41: 226–31.
13. Huang HC, Jiang ZF (2009) Accumulated amyloid-beta peptide and hyperphosphorylated tau protein: relationship and links in Alzheimer's disease. J Alzheimers Dis 16: 15–27.
14. Jin M, Shepardson N, Yang T, Chen G, Walsh D, et al. (2011) Soluble amyloid beta-protein dimers isolated from Alzheimer cortex directly induce Tau hyperphosphorylation and neuritic degeneration. Proc Natl Acad Sci U S A 108: 5819–24.
15. Lin H, Bhatia R, Lal R (2001) Amyloid β protein forms ion channels: implications for Alzheimer's disease pathophysiology. FASEB J 15: 2433–2444.
16. Quist A, Doudevski I, Lin H, Azimova R, Ng D, et al. (2005) Amyloid ion channels: a common structural link for proteinmisfolding disease. Proc Natl Acad Sci USA 102: 10427–10432.
17. Rhee SK, Quist AP, Lal R (1998) Amyloid β protein-(1–42) forms calcium-permeable, Zn2+-sensitive channel. J Biol Chem 273: 13379–13382.
18. Jang H, Arce FT, Ramachandran S, Capone R, Azimova R, et al. (2010) Truncated beta-amyloid peptide channels provide an alternative mechanism for Alzheimer's Disease and Down syndrome. Proc Natl Acad Sci U S A 107: 6538–43.
19. Zádori Z, Szelei J, Lacoste M, Li Y, Gariépy S, et al. (2001) A Viral Phospholipase A2 Is Required for Parvovirus Infectivity. Developmental Cell 1: 291–302.
20. Aravind L, Koonin E (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem Sci 23: 469–472.
21. Gomis-Rüth FX, Kress LF, Kellermann J, Mayr I, Lee X, et al. (1994) Refined 2.0 A X-ray crystal structure of the snake venom zinc-endopeptidase adamalysin II. Primary and tertiary structure determination, refinement, molecular structure and comparison with astacin, collagenase and thermolysin. J Mol Biol 239: 513–44.
22. Thöny-Meyer L (1997) Biogenesis of respiratory cytochromes in bacteria. Microbiol Mol Biol Rev 61: 337–376.
23. Dennis EA (1997) The growing phospholipase A2 superfamily of signal transduction enzymes. Trends Biochem Sci 22: 1–2.
24. Murakami M, Nakatani Y, Atsumi G, Inoue K, Kudo I (1997) Regulatory functions of phospholipase A2. Crit Rev Immunol 17: 225–283.
25. Dessen A (2000) Phospholipase A(2) enzymes: structural diversity in lipid messenger metabolism. Structure 8: R15–R22.
26. Zimmerman MD, Proudfoot M, Yakunin A, Minor W (2008) Structural insight into the mechanism of substrate specificity and catalytic activity of an HD-domain phosphohydrolase: the 5′-deoxyribonucleotidase YfbR from Escherichia coli. J Mol Biol 378: 215–26.
27. Yakunin AF, Proudfoot M, Kuznetsova E, Savchenko A, Brown G, et al. (2004) The HD domain of the Escherichia coli tRNA nucleotidyltransferase has 2′,3′-cyclic phosphodiesterase, 2′-nucleotidase, and phosphatase activities. J Biol Chem 279: 36819–27.
28. Maskos K, Fernandez-Catalan C, Huber R, Bourenkov GP, Bartunik H, et al. (1998) Crystal structure of the catalytic domain of human tumor necrosis factor-alpha-converting enzyme. Proc Natl Acad Sci U S A 95: 3408–12.
29. Schulz H, Flennecke H, Thöny-Meyer L (1998) Prototype of a heme chaperone essential for cytochrome c maturation. Science 281: 1197–1200.
30. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16: 10881–90.
31. Tusnády GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics 17: 849–50.
32. Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.
33. Holland B, Moulton V (2003) Consensus Networks: A Method for Visualising Incompatibilities in Collections of Trees. Lecture Notes in Bioinformatics 2812: 165–176.
34. Huson DH, Bryant D (2006) Application of Phylogenetic Networks in Evolutionary Studies. Mol Biol Evol 23: 254–267.
35. Rivas E (2005) Evolutionary models for insertions and deletions in a probabilistic modeling framework. BMC Bioinformatics 6: 63.
36. Felsenstein J (2003) Inferring Phylogenies. SunderlandMassachusetts: Sinauet Associates. pp 204–206.
37. Baldi P, Brunak S (2001) Bioinformatics. London: The MIT Press. pp 267–269.
38. Durbin R, Eddy S, Krogh A, Mitchison G (1998) Biological sequence analysis. Cambridge: Cambridge Univerity Press. pp 196–197.
39. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, et al. (2010) Protein annotation and modelling servers at University College London. Nucleic Acids Res 38(Web Server issue): W563–8.
40. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36(Web Server issue): W197–201.
41. Cai CZ, Han LY, Ji ZL, Chen X, Chen YZ (2003) SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence. Nucleic Acids Res 31: 3692–7.

42. NaiK PK, Ranjan P, Kesari P, Jain S (2011) MetalloPred: a tool for hierarchical prediction of metal ion binding proteins using cluster of neural networks and sequence derived features. J Biophys Chem 2: 112–123.

43. Bayer TA, Paliga K, Weggen S, Wiestler OD, Beyreuther K, et al. (1997) Amyloid precursor-like protein 1 accumulates in neuritic plaques in Alzheimer's disease. Acta Neuropathol 94: 519–24.

44. Iijima K, Lee DS, Okutsu J, Tomita S, Hirashima N, et al. (1998) cDNA isolation of Alzheimer's amyloid precursor protein from cholinergic nerve terminals of the electric organ of the electric ray. Biochem J 330: 29–33.

45. Villard L, Tassone F, Crnogorac-Jurceviæ T, Clancy K, Gardiner K (1998) Analysis of pufferfish homologues of the AT-rich human APP gene. Gene 210: 17–24.

46. Ghiso J, Rostagno A, Gardella JE, Liem L, Gorevic PD, et al. (1992) A 109-amino-acid C-terminalf ragment of Alzheimer's-diseasae amyloid precursor protein contains a sequence, -RHDS- that promotes cell adhesion. Biochem J 288: 1053–9.

47. Sabo S, Lambert MP, Kessey K, Wade W, Krafft G, et al. (1995) Interaction of beta-amyloid peptides with integrins in a human nerve cell line. Neurosci Lett 184: 25–28.

48. Matter ML, Zhang Z, Nordstedt C, Ruoslahti E (1998) The alpha5beta1 integrin mediates elimination of amyloid-beta peptide and protects against apoptosis. J Cell Biol 141: 1019–30.

49. Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, et al. (2002) Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. Science 296: 151–155.

50. Xiao T, Takagi J, Coller BS, Wang JH, Springer TA (2004) Structural basis for allostery in integrins and binding to fibrinogenmimetic therapeutics. Nature 432: 59–67.

51. Takagi J (2007) Structural basis for ligand recognition by integrins. Curr Opin Cell Biol 19: 557–64.

52. Friedberg I, Margalit H (2002) Persistently conserved positions in structurally similar, sequence dissimilar proteins: roles in preserving protein fold and function. Protein Sci 11: 350–60.

53. Zhang Y, Stec B, Godzik A (2007) Between order and disorder in protein structures:analysis of "dual personality" fragments in proteins. Structure 15: 1141–7.

54. Kabsch W, Sander C (1984) On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. Proc Natl Acad Sci U S A 81: 1075–8.

55. Krishna N, Guruprasad K (2011) Certain heptapeptide and large sequences representing an entire helix, strand or coil conformation in proteins are associated as chameleon sequences. Int J Biol Macromol 49: 218–22.

56. Zirah S, Kozin SA, Mazur AK, Blond A, Cheminant M, et al. (2006) Structural changes of region 1–16 of the Alzheimer disease amyloid beta-peptide upon zinc binding and in vitro aging. J Biol Chem 281: 2151–61.

57. Barrow CJ, Zagorski MG (1991) Solution structures of beta peptide and its constituent fragments: relation to amyloid deposition. Science 253: 179–82.

58. Zagorski MG, Barrow CJ (1992) NMR studies of amyloid beta-peptides: proton assignments, secondary structure, and mechanism of an alpha-helix----beta-sheet conversion for a homologous, 28-residue, N-terminal fragment. Biochemistry 31: 5621–31.

59. Crescenzi O, Tomaselli S, Guerrini R, Salvadori S, D'Ursi AM, et al. (2002) Solution structure of the Alzheimer amyloid beta-peptide (1–42) in an apolar microenvironment. Similarity with a virus fusion domain. Eur J Biochem 269: 5642–8.

60. Coles M, Bicknell W, Watson AA, Fairlie DP, Craik DJ (1998) Solution structure of amyloid beta-peptide(1–40) in a water-micelle environment. Is the membrane-spanning domain where we think it is? Biochemistry 37: 11064–77.

61. Sticht H, Bayer P, Willbold D, Dames S, Hilbich C, et al. (1995) Structure of amyloid A4-(1–40)-peptide of Alzheimer's disease. Eur J Biochem 233: 293–8.

62. Lührs T, Ritter C, Adrian M, Riek-Loher D, Bohrmann B, et al. (2005) 3D structure of Alzheimer's amyloid-beta(1–42) fibrils. Proc Natl Acad Sci U S A 102: 17342–7.

63. Päiviö A, Nordling E, Kallberg Y, Thyberg J, Johansson J (2004) Stabilization of discordant helices in amyloid fibril-forming proteins. Protein Sci 13: 1251–9.

64. Long JA, Gordon MS (2004) The greatest step in vertebrate history: a paleobiological review of the fish-tetrapod transition. Physiol Biochem Zool 77: 700–19.

65. Regier JC, Shultz JW, Kambic RE (2005) Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc Biol Sci 272: 395–401.

66. Burmester T (2001) Molecular evolution of the arthropod hemocyanin superfamily. Mol Biol Evol 18: 184–95.

67. Levin I, Miller MD, Schwarzenbacher R, McMullan D, Abdubek P, et al. (2005) Crystal structure of an indigoidine synthase A (IndA)-like protein (TM1464) from Thermotoga maritime at 1.90 A resolution reveals a new fold. Proteins 59: 864–8.

68. Liu J, Lou Y, Yokota H, Adams PD, Kim R, et al. (2005) Crystal structure of a PhoU protein homologue: a new class of metalloprotein containing multinuclear iron clusters. J Biol Chem 280: 15960–6.

69. Wawrzak Z, Sandalova T, Steffens JJ, Basarab GS, Lundgvist T, et al. (1999) High-resolution structures of scytalone dehydratase-inhibitor complexes crystal-lized at physiological pH. Proteins 35: 425–39.

70. Jenkins J, Shevchik VE, Hugouvieux-Cotte-Pattat N, Pickersgill RW (2004) The crystal structure of pectate lyase Pel9A from Erwinia chrysanthemi. J Biol Chem 279: 9139–45.

71. Schumacher MA, Carter D, Ross DS, Ullman B, Brennan RG (1996) Crystal structures of Toxoplasma gondii HGXPRTase reveal the catalytic role of a long flexible loop. Nat Struct Biol 3: 881–7.

72. Chander P, Halbig KM, Miller JK, Fields CJ, Bonner HK, et al. (2005) Structure of the nucleotide complex of PyrR, the pyr attenuation protein from Bacillus caldolyticus, suggests dual regulation by pyrimidine and purine nucleotides. J Bacteriol 187: 1773–82.

73. Yang Z, Zhang H, Hung HC, Kuo CC, Tsai LC, et al. (2002) Structural studies of the pigeon cytosolic NADP(+)-dependent malic enzyme. Protein Sci 11: 332–41.

74. Zuo Y, Vincent HA, Zhang J, Wang Y, Deutscher MP, et al. (2006) Structural basis for processivity and single-strand specificity of RNase II. Mol Cell 24: 149–56.

75. Harding MM (2004) The architecture of metal coordination groups in proteins. Acta Crystallogr D Biol Crystallogr 60: 849–59.

76. Lovell MA, Robertson JD, Teesdale WJ, Campbell JL, Markesbery WR (1998) Copper, iron and zinc in Alzheimer's disease senile plaques. J Neurol Sci 158: 47–52.

77. Dong J, Atwood CS, Anderson VE, Siedlak SL, Smith MA, et al. (2003) Metal-binding and oxidation of amyloid-beta within isolated senile plaque cores: Raman microscopic evidence. Biochemistry 42: 2768–73.

78. Minicozzi V, Stellato F, Comai M, Dalla Serra M, Potrich C, et al. (2008) Identifying the minimal copper- and zinc-binding site sequence in amyloid-beta peptides. J Biol Chem 283: 10784–92.

79. Parthasarathy S, Long F, Miller Y, Xiao Y, McElheny D, et al. (2011) Molecular-level examination of Cu2+ binding structure for amyloid fibrils of 40-residue Alzheimer's β by solid-state NMR spectroscopy. J Am Chem Soc 133: 3390–400.

80. Gaggelli E, Janicka-Klos A, Jankowska E, Kozlowski H, Migliorini C, et al. (2008) NMR studies of the Zn2+ interactions with rat and human beta-amyloid (1–28) peptides in water-micelle environment. J Phys Chem B 112: 100–109.

81. Roher AE, Lowenson JD, Clarke S, Wolkow C, Wang R, et al. (1993) Structural alterations in the peptide backbone of beta-amyloid core protein may account for its deposition and stability in Alzheimer's disease. J Biol Chem 268: 3072–83.

82. Miller Y, Ma B, Nussinov R (2010) Zinc ions promote Alzheimer Abeta aggregation via population shift of polymorphic states. Proc Natl Acad Sci U S A 107: 9490–5.

83. Sarell CJ, Syme CD, Rigby SE, Viles JH (2009) Copper(II) binding to amyloid-beta fibrils of Alzheimer's disease reveals a picomolar affinity: stoichiometry and coordination geometry are independent of Abeta oligomeric form. Biochemistry 48: 4388–402.

84. Hancock JM, Shaw PJ, Bonneton F, Dover GA (1999) High sequence turnover in the regulatory regions of the developmental gene hunchback in insects. Mol Biol Evol 16: 253–265.

85. Moses AM, Pollard DA, Nix DA, Iyer VN, Li XY, et al. (2006) Large-scale turnover of functional transcription factor binding sites in Drosophila. PLoS Comput Biol 2: e130.

86. Venkataram S, Fay JC (2010) Is transcription factor binding site turnover a sufficient explanation for cis-regulatory sequence divergence? Genome Biol Evol 2: 851–8.

87. Lim KH, Kim YK, Chang YT (2007) Investigations of the molecular mechanism of metal-induced Abeta (1–40) amyloidogenesis. Biochemistry 46: 13523–32.

88. Shaked GM, Kummer MP, Lu DC, Galvan V, Bredesen DE, et al. (2006) Abeta induces cell death by direct interaction with its cognate extracellular domain on APP (APP 597–624). FASEB J 20: 1254–6.

89. Cao X, Südhof TC (2001) A transcriptionally [correction of transcriptively] active complex of APP with Fe65 and histone acetyltransferase Tip60. Science 293: 115–20.

90. Ghersi E, Noviello C, D'Adamio L (2004) Amyloid-beta protein precursor (AbetaPP)intracellular domain-associated protein-1 proteins bind to AbetaPP and modulate its processing in an isoform-specific manner. J Biol Chem 279: 49105–12.

91. Osterfield M, Egelund R, Young LM, Flanagan JG (2008) Interaction of amyloid precursor protein with contactins and NgCAM in the retinotectal system. Development 135: 1189–99.

92. King GD, Perez RG, Steinhilb ML, Gaut JR, Turner RS (2003) X11alpha modulates secretory and endocytic trafficking and metabolism of Amyloid Precursor Protein: mutational analysis of the YENPTY sequence. Neuroscience 120: 143–154.

93. Hao CY, Perkinton MS, Chan WW, Chan HY, Miller CC, et al. (2011) GULP1 is a novel APP-interacting protein that alters APP processing. Biochem J 436: 631–9.

94. Wakutani Y, Watanabe K, Adachi Y, Wada-Isoe K, Urakami H, et al. (2004) Novel amyloid precursor protein gene missense mutation (D678N) in probable familial Alzheimer's disease. J Neurol Neurosurg Psychiatry 75: 1039–1042.

95. Janssen JC, Beck JA, Campbell TA, Dickinson A, Fox NC, et al. (2003) Early onset familial Alzheimer's disease: Mutation frequency in 31 families. Neurology 60: 235–239.