Chapter 5

Experimental Molecular Archeology: Reconstruction of Ancestral Mutants and Evolutionary History of Proteins as a New Approach in Protein Engineering

Tomohisa Ogawa and Tsuyoshi Shirai

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/56255

1. Introduction

The diversity of life on Earth is the result of perpetual evolutionary processes beginning at life's origins; evolution is the fundamental development strategy of life. Today, studies of gene and protein sequences, including various genome-sequencing projects, provide insight into these evolutionary processes and events. However, the sequence data obtained is restricted to extant genes and proteins, with the exception of the rare fossil genome samples [1, 2], for example Neanderthal [3], archaic hominin in Siberia [4, 5], and ancient elephants such as mastodon and mammoth [6]. The fossil record, and genome sequences derived from it, has the potential to elucidate ancient, extinct forms of life, acting as missing links to fill evolutionary gaps; however, the sequenced fossil genome is very limited, mainly due to the condition of samples and the challenges of preparing them. Discovering the forms of ancient organisms is one of the major purposes of paleontology, and is valuable in understanding of current life forms as these will be a reflection of their evolutionary history. However, the reconstruction of a living organism from fossils, which would be the ultimate paleontological methodology, is far beyond the currently available technologies, although there has recently been a report of the production of an artificial bacterial cell, using a chemically synthesized genome [7].

Meanwhile, for genes or the proteins they encode, it is already feasible to reconstruct their ancestral forms using phylogenetic trees constructed from sequence data; these techniques may well provide clues to the evolutionary history of certain extant genes and proteins with respect to their ancestors. Although phylogenetic analyses alone, or in combination with protein structure simulations, are useful to analyze structure-function relationships and evolutionary history [8], resurrected ancient recombinant proteins have the potential to

provide more direct observations. Production of ancestral or ancient proteins can be achieved comparably easily due to developments in molecular biology and protein engineering techniques, which allow nucleotide or amino acid sequences to be synthesized. Ancestral proteins can be tested in the laboratory using biochemical or biophysical methods, for their activity, stability, specificity, and even three-dimensional structure. Thus, ancestral sequence reconstruction (ASR) has proved a useful experimental tool for studying the diverse structure and function of proteins [9]. To date, such 'experimental molecular archeology' using ASR has been applied to several enzymes [10-24], including photo-reactive proteins [25-37], nuclear receptor and transmembrane proteins [38-48], lectins [49-52], viral proteins [53, 54], elongation factor [55-57], paralbmin [58], in addition to a number of peptides [59,60] (Table 1).

In early studies, ASR experiments using the technique of molecular phylogeny were based on basic site-directed mutagenesis and used to investigate the functional evolution of proteins, including the convergent evolution of lysozyme in ruminant stomach environments and the adaptation of enzymes to alkaline conditions [10-12]. However, if ancestral sequences have been determined, the most straightforward method is to reconstruct the full-length ancestral protein in the laboratory. No fundamental differences exist between ancestor reconstruction and standard site-directed mutagenesis, other than the number of amino acids residues requiring mutation, which, in the case of ancestor reconstruction, might be spread over the entire sequence. At present, ASR can be achieved using commercially available *de novo* synthetic genes. Thus, 'experimental molecular archeology' by ancestral protein reconstruction using a combination of the technical developments in biochemistry, molecular biology, and bioinformatics can be exploited in both molecular evolutionary biology and protein engineering. In this chapter, we will provide an overview the experimental molecular archeology technique of ASR, and the case of ancestral fish galectins will be discussed in detail, based on our recent studies.

2. The early studies: Reconstruction of partial ancestors by site-directed mutagenesis

The first studies exploiting the idea of ancestral protein reconstruction used site-direct mutagenesis, in which a small number of amino acids were substituted to produce the anticipated ancestral status. These studies include the reconstruction of a ribonuclease (RNase) of an extinct bovid ruminant [10, 11], and the lysozymes from a game-bird using ancestral lysozyme reconstructions predicted by the MP (Maximum Parsimony) method [12]. Benner and colleagues reconstructed RNase of an extinct bovid ruminant [10], by predicting four sequences of ancestral RNases from five closely related bovids including ox, swamp buffalo, river buffalo, nilgai, and the primitive artiodactyl using the MP method [61, 62]. The ancestor closest to the extant ox protein was selected from the four probable ancestors as the target of the experiment as it contained a mutation of amino acid residue 35, located close to Lys41, which is known to be important for catalysis. Three ancestral mutants of the ox RNase (A19S, L35M, and A19S/L35M) were examined for their kinetic properties and the thermal stabilities against tryptic digestion. However, no significant difference was found between the ancestral

http://dx.doi.org/10.5772/56255

Ancestral proteins (family)	Methods/programs	Partial/ full length	Remarkable future	References
[Enzymes]				
RNases		point mutants		[10, 11]
Lysozyme				[12]
Alchole dehydrogenase	ML (PAML)	full length	similar to Adh1 than Adh2 accumulation of ethanol	[13]
Isopropylmalate dehydrogenase (IPMDH)		multiple-point mutants		[14-16]
Isocitrate dehydrogenase (ICDH)				[17] [18]
3-isopropylmalate dehydrogenase (LeuB)	ML/Bayesian	full length	ancestral 3D structures	[19]
Chymase	MP	full length	ancestor of ◎/◎-chymase Ang II-forming activity	[20]
DNA gyrase	ML (Tree-Puzzle)	Partial: ATPase D	thermal stability	[21]
Glycyl-tRNA synthetase (GlyRS)	ML	multiple-point mut	Commonote, thermal stability	[22]
Sulfotransferases /Paraoxonase	ML (FastML)	multiple-point mut	Directed evolution ancestral 3D structure	[23]
Thioredoxin (Trx)			Precambrian enzyme (anoxygenic/oxygenic environment)	[24]
[Visual pigment proteins & Fluorescent proteins]				
Opsins (rhodopsin)	ML (nucleotide/amino acids/codon)		Archosaur ancestor	[25]
		multiple-point mut	UV pigment (SWS1)	[26]
	ML-based Bayesian	multiple-point mut	Rhodopsin (RH1)	[27]
		multiple- point mut	Red/green color vison	[28]
			Zebrafish RH2-1~4	[29]
	ML-based Bayesian (PAML), MP	multiple-point mut	Dim-light (deep-sea) vision	[30-32]
GFP (coral pigment)	ML (MrBayes 3.0)	full length	Fluorescence spectra evolved from green common ancestor convergent evolution/positive selection	[33-37]
[Receptors & transmembrane proteins]				
Nuclear receptors (NR) for steroid estrogen/androgen/progesterone/ glucocorticoid/mineralocorticoid receptors	MP/ML		ancestral 3D structures NR superfamily	[38-47]
Vacuolar H1-ATPase				[48]
[Carbohydrate binding proteins/Lectin]				
Galectins	ML/MP	full length	carbohydrate binding ancestral 3D structures	[49-51]
Tachlectin-2 (◎-propeller lectin)		fragments	oligomeric assembly	[52]
[Viral proteins]				
Coxsackievirus B5 capsid	ML (PAML)	P1 region	Infectious activity, cell binding Cell tropism, antigenicity	[53]
Core protein PtERV1 p12-Capsid	MP	point mut	TRIM5α antiviral protein	[54]
[Others]				
Elongation factors (EF) Tu	ML (MOLPHY/PAML/ JTT/Dayhoff/WAG)	full length	thermostability [55-57] phenotype genotype Hyperthermophiles	
Parvalbumin (PVs)	ML (FastML)	point mutants	thermal adaptation	[58]
Allatostatins (ASTs)	ML (PAML)/ consider gap	peptide	juvenile hormone release inhibition	[59]
Glucagon-like peptide-1 (GLP-1)	ML (FastML) with JTTmatrix	peptide receptor	affinity, stability	[60]

ML: Maximum likelihood/bayesian, MP: Maximum Parsimony

Table 1. The experimental molecular archeology analysis using ancestral proteins

mutants and the modern ox RNase. The results suggested that these amino acid substitutions were evolutionarily neutral, although this conclusion is limited to the extent of the examined properties [11].

Malcolm et al. succeeded in identifying a non-neutral evolutionary pathway of game-bird lysozymes using ancestral lysozyme reconstructions predicted by the MP method [12]. Seven mutations in game-bird lysozyme proteins included combinations of residues Thr40, Ile55, and Ser91, which were anticipated to be Ser40, Val55 and Thr91, respectively, in ancestral molecules. The mutants were synthesized as possible intermediates in the evolutionary pathway of bird lysozyme and comparative molecular properties and crystal structures of these revealed that the thermostabilities of the proteins were correlated with the bulkiness of their side chains. The T40S mutant increased its thermostability by more than 3°C, allowing the conclusion that this mutation was non-neutral effect of natural selection.

Yamagishi and colleagues used ancestral protein reconstruction [14-16] to obtain direct evidence for the hypothesis that the common ancestor of all organisms was hyper-thermophilic [63]. Because the catalytic activities of 3-isopropylmalate dehydrogenase (IPMDH) and isocitrate dehydrogenase (ICDH) are similar to one another and their three-dimensional structures conserved, these proteins are diverged from an ancient common ancestor [64], of which sequence was inferred from a phylogenetic tree constructed from IPMDH and ICDH sequences from various species, including the thermophile (*Thermus thermophilus*) and the extreme thermophile (*Sulfolobus* sp. strain 7). Five of the seven ancestral mutants, in which substituted amino acids were located close to the substrate and cofactor-binding sites, demonstrated higher thermostability than wild type IPMDH from *Sulfolobus* sp. strain 7. These findings were taken to support the hypothesis of a hyperthermophile common ancestor. Moreover, the successful thermostabilization of ICDH [17] and Glycyl-tRNA synthetase [22] by ASR has been reported. Thus, the incorporation of ancestral residues into a modern protein can be used not only to test evolutionary hypotheses, but also as a powerful protein engineering technique for protein thermostabilization.

Recently, Whittington and Moerland reported that ASR analysis of parvalbumins (PVs) was able to identify the set of substitutions most likely to have caused a significant shift in PV function during the evolution of *Antarctic notothenioids* in the frigid waters of the Southern Ocean [58]. The results suggest that the current thermal phenotype of Antarctic PVs can be recapitulated by only two amino acid substitutions, namely, K8N and K26N.

These studies were performed by introducing a limited number of mutations into extant proteins, or by carefully selecting ancestors that were separated from an extant protein by only few substitutions. However, such ancestral reconstruction by site-directed mutagenesis appears to be incomplete, as the possibility that sites remaining in a non-ancestral state may significantly affect the molecular property of interest, cannot be ruled out. Although it is difficult and expensive to introduce many mutations into sites widely distributed over gene sequences by site-directed mutagenesis, *de novo* gene synthesis is now available, allowing preparation of ancestral proteins. Therefore, the majority of recent ASR studies have been conducted using full-length or partial ancestral sequence reconstruction, including substitution of corresponding sites in target proteins.

http://dx.doi.org/10.5772/56255

3. Methods for ancestral sequence prediction

How can we determine the sequences of ancestral proteins or genes? In most cases, since the ancestral genes do not currently exist, the ancestral sequences need to be estimated and reconstructed mainly *in silico* (using a computer). Ancestral sequences are calculated using computational methods originally developed for molecular phylogeny construction. Some of these methods, such as maximum parsimony (MP) and maximum likelihood (ML), have an integral procedure of ancestral sequence inference at each node of the phylogenetic tree under construction [65, 66]. The MP method assumes that a phylogenetic tree with minimum substitutions is the most likely. This method assigns a possible nucleotide/amino acid for each site at every node of a phylogenetic tree to evaluate the minimum substitution number. Because of this assumption of parsimony, the MP method tends to underestimate the number of substitutions if a branch is relatively long. The method is also fragile if the evolutionary rate varies among branches.

By contrast, the ML method, which does not require this assumption, is currently more widely used. This method evaluates the posterior probability of a nucleotide/amino acid residue at each node of a phylogenetic tree, based on empirical Bayesian statistics, using the provided sequences and a substitution probability matrix as inputs (observations). Therefore, results can be significantly affected by the choice of input sequences and the choice of substitution probability matrix; the probability of a reconstructed sequence at a node might be low when the node is connected to the provided sequences through longer and/or more intervening branches. The ML method is popular in the field, largely owing to the presence of the excellent software package PAML [67]. Several other software applications have been also developed for this purpose, such as FastML [68], ANSESCON [69], and GASP [70]. With the exception of GASP which partly employs the MP method to enable ancestral state prediction at gapped sites in a sequence alignment, these applications are based on the ML method. In many cases, ancestral sequences cannot be unambiguously determined, and several amino acids might be assigned to a residue site with almost equal probabilities. To avoid false conclusions as a result of such ambiguity, the accuracy of reconstructed ancestral sequence is critical for such studies. However, it is often difficult to obtain a complete, highly accurate sequence, as molecular evolution is believed to be a highly stochastic process and there is no guarantee that ancestral sequences can be identified without errors. Even if each residue of a protein made up of 100 residues, is identified with posterior probability of 0.99 (ie. 99% are expected to be correct), the probability that the sequence as a whole is accurate is only ~0.37 (i.e., 0.99^{100}). In many actual cases, site probabilities are likely to be much lower. This is a major problem in ancestor reconstruction studies, and considerable efforts have been made to avoid incorrect conclusions due to imperfect reconstruction.

Williams et al reported the assessment of the accuracy of ancestral protein reconstruction by MP, ML and Bayesian inference (BI) methods [71]. Their results indicated that MP and ML methods, which reconstruct "best guess" amino acids at each position, overestimate thermostability, while the BI method, which sometimes chooses less-probable residues from the posterior probability distribution, does not. ML and MP tend to eliminate variants at a position that are slightly detrimental to structural stability, simply because such detrimental variants are less frequent. Thus, Williams et al caution that ancestral reconstruction studies

require greater care to come to credible conclusions regarding functional evolution [71]. Thornton and colleagues also examined simulation-based experiments, under both simplified and empirically derived conditions, to compare the accuracy of ASR carried out using ML and Bayesian approaches [72]. They showed that incorporating phylogenetic uncertainty by integrating over topologies very rarely changes the inferred ancestral state and does not improve the accuracy of the reconstructed ancestral sequence, suggesting that ML can produce accurate ASRs, even in the face of phylogenetic uncertainty, and using Bayesian integration to incorporate the uncertainty is neither necessary nor beneficial [72].

In the case for experimental molecular archeology using ASR, the effects of equally probable residues at unreliable sites have been tested by site-directed mutagenesis to confirm directly that molecular properties are not largely affected by these. Indeed, in the case of ancestral congerin genes, the single mutant Con-anc'-N28K, in which the suspicious site was replaced with the alternate suggested amino acid was reconstructed in addition to the ancestral congerin (Con-anc', the last common ancestor of ConI and ConII) inferred from the phylogeny of extant galectins using the ML method based on DNA sequences [51]. Nucleotide sequences were retrieved from the DDBJ database [73] , and the ancestral sequence were inferred using the PAML program [67]. The alignment of amino acid sequences of the extant galectins was first prepared using the XCED program [74], and an alignment of the corresponding nucleotide sequences was made in accordance with the amino acid sequence alignment. Tree topology was based on the amino acid sequences of extant proteins using the neighbor-joining (NJ) method. PAML was applied to the phylogeny and alignment to infer the ancestral sequences. The F1X4 matrix was used as the codon substitution model with the universal codon table. The free *dN*/*dS* ratio with M8 (beta & omega) model was adapted [75]. The reproduction rate of each Con-anc' amino acid residue was also calculated from the reconstructed sequences, with the exclusion of one extant gene in each case, in order to identify highly unstable sites depending on the choice of extant genes. The results indicated that the average reproduction rate over the sequence was 0.98. The average site posterior probability in the sequence of Con-anc' was 0.81. Seventy-two of 135 sites (53%) had a posterior probability > 0.9. By contrast, 11 sites were found to have posterior probabilities < 0.5. Only one residue, Asn28 of Con-anc', was reproduced with a distinguishably low rate of 0.286, with a suggested alternative amino acid of Lys. Therefore, the single mutant Con-anc'-N28K was also reconstructed. Several reconstruction tests demonstrated that the ancestral sequence had constantly converged into that of Con-anc', and the expected shift by adding a newly found extent sequence was reduced to 1.4% (s.d. 3.2%).

In the case of alcohol dehydrogenase (Adh) ancestral mutants reported by Thomson et al., the posterior probability of the sequence predicted by the ML method was found to be low at three sites. Amino acid residues 168, 211 and 236 of Adh had two (Met and Arg), three (Lys, Arg and Thr), and two (Asp and Asn) equally probable candidates as the ancestral residues, respectively. Therefore, all possible combinations (2 x 3 x 2 = 12) of the candidates at the ambiguous sites were reproduced, and their kinetic properties assessed [13]. The results confirmed with consistency among the alternative mutants that acetaldehyde metabolism was the original function of Adh, that ancestral yeast could not consume ethanol, and that the function of ethanol metabolism was most likely acquired in the linage of the Adh2 locus after gene duplication.

http://dx.doi.org/10.5772/56255

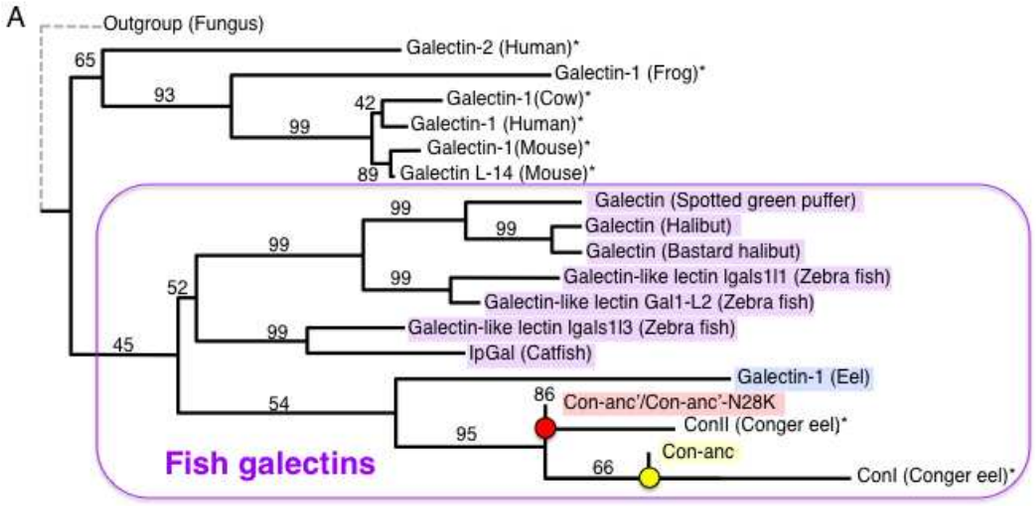
4. Reconstruction of full-length ancestral proteins: Selective adaptive evolution of Conger eel galectins

Conger eel galectins, termed Congerins I and II (Con I and Con II), function as biodefense molecules in the skin mucus and frontier organs including the epidermal club cells of the skin, wall of the oral cavity, pharynx, esophagus, and gills [76-79]. Con I and Con II are prototype galectins, composed of subunits containing 135 and 136 amino acids, respectively, and display 48% amino acid sequence identity [80]. While both Con I and Con II form 2-fold symmetric homodimers with 5- and 6-stranded β-sheets (termed a jellyroll motif), they have different stabilities and carbohydrate-binding specificities, although they do have the conserved carbohydrate recognition domain (CRD) common to other galectins [81-84]. Previous studies of Con I and Con II, based on molecular evolutionary and X-ray crystallography analyses, revealed that these proteins have evolved via accelerated substitutions under natural selection pressure [74-85].

To understand the rapid adaptive differentiation of congerins, experimental molecular archaeology analysis, using the reconstructed ancestral congerins, Con-anc and Con-anc', and their mutants has been conducted [49-51]. Since the ancestral sequences of congerin, Con-anc and Con-anc', were estimated from different phylogenetic trees, which were constructed from the varying numbers of extant genes available (eight for Con-anc, and sixteen for Con-anc') (Fig. 1A), the ancestral sequence Con-anc' showed a 27% discrepancy from the previously inferred sequence of Con-anc (Fig. 1B). Furthermore, as described in the 'Methods for Ancestral Sequence Prediction' section, the reproduction rate of each Con-anc' amino acid residue was examined for the reconstructed sequences, with one extant gene excluded for each estimation, in order to identify highly unstable sites. The result indicated that the average reproduction rate over the sequences were 0.98, and only one residue, Asn28 of Con-anc', was reproduced with a distinguishably low rate of 0.286, prompting verification of the results by the construction of a single mutant Con-anc'-N28K. The revised ancestral congerins, Con-anc' or Con-anc'-N28K, were attached to the nodes of extant proteins with zero distance in the phylogeny constructed from amino acid sequences, indicating that the sequence was appropriate for that of an ancestor (Fig. 1A). On the other hand, the previously inferred Con-anc was attached midway on the ConI branch. Therefore, Con-anc' or Con-anc'-N28K are likely to be closer to the true common ancestor of ConI and ConII than Con-anc. The structures and molecular properties of congerins, as discussed below, also supported this conclusion.

Although Con-anc is an ancestral mutant located midway on the ConI branch and shares a higher sequence similarity with ConI (76%) than with ConII (61%), it showed unique carbohydrate-binding activity and properties, and more closely resembled ConII than ConI, in terms of thermostability and carbohydrate recognition specificity, with the exception of carbohydrates containing α2, 3-sialyl galactose, for example GM3 and GD1a. The ancestral congerins, Con-anc' and Con-anc'-N28K, demonstrated similar carbohydrate binding activity and specificities to those of Con-anc [51]. These analyses of Con-anc suggested a functional evolutionary process for ConI, where it evolved from the ancestral congerin to increase its structural stability and sugar-binding activity. In the case of the ancestral congerin, Con-anc,

the candidate amino acid residues responsible for the higher structural stability and carbohydrate-binding activity of Con I were reduced to only 31 amino acid residues, from a total of 71 with apparent differences between Con I and Con II. These were mainly located in the N- and C-terminal and loop regions of the molecule, including the CRD [49, 50]. To identify the residues responsible for the properties of Con I, we next performed molecular evolution tracing analysis, by constructing pseudo-ancestral Con-anc proteins focused on the N-terminal, C-terminal, and some loop regions (loops 3, 5 and 6) [50].



B

```
Con-anc'          MSDGVEVKNITFKLGMYLTVGGVVNSNANRFSINVGESTDSIALHIDHRFSYGADQNTIVLNSMVDD--  67
Con-anc'-N28K     ..........................K.....................................--  67
Con-anc           ..G.L...FD..V.KF......I.NY.K..........N.LS..L....N.............NS--  67
ConI              ..G.LQ...FD.TV.KF.....FI.NSPQ...V.....MN.LS..L....N.........M..TLKGDN  69
ConII             ...RA..R..P.................T.............M.M...........VL....L.HNV-  68

Con-anc'          GWQQEQRSKNFPFTAGEHFQITITFDIDTFYIQLPDGHKVEFPNRHGDEAFNFIYLAGDARLTFVRLE  135
Con-anc'-N28K     ........................................................................  135
Con-anc           ..ET.........S...Y.E......TNK...E.L....L.....YKK..L..LS..........K..  135
ConI              ..ET....T..TLS..QY.E..LSY.INK...DIL..PNL.....YSK.FLP.LS........L.K..  137
ConII             .....E...K....K.D...T.....TH......SN.ET......NK.A...L..............  136
```

(A) Phylogeny of extant and ancestral congerins. The tree is based on the amino acid sequences of extant galectins and ancestral congerins. The extant genes used for ancestral reconstruction and their accession codes are ConI (*Conger myriaster* congerin I, AB010276.1), ConII (*C. myriaster* congerin II, AB010277.1), *Anguilla japonica* galectin-1 (AJL1, AB098064.1), *Hippoglossus hippoglossus* galectin (AHA1, DQ993254.1), *Paralichthys olivaceus* galectin (PoGal, AF220550.1), *Tetraodon nigroviridis* galectin (TnGal, CR649222.2), *Danio rerio* galectin-like lectin lgals1l1 (DrGal1_L1, BC164225.1), *D. rerio* galectin-like lectin Gal1-L2 (DrGal1_L2, AY421704.1), *D. rerio* Galectin-like lectin lgals1l3 (DrGal1_L3, BC165230.1), *Ictalurus punctatus* galectin (IpGal, CF261531), *Bos taurus* galectin-1 (BTG1, BC103156.1), *Homo sapiens* galectin-1 (HSG1, AK312161.1), *Mus musculus* galectin-1 (MMG1, BC099479.1), *Cricetulus* sp. galectin L-14 (CRG1, M96676.1), *Xenopus laevis* galectin-1 (XLG1, AF170341.1), and *H. sapiens* galectin-2 (HSG2, BC059782.1). The numbers associated with the branches are the percent reproductions of branches in 1000 bootstrap reconstructions. This tree is rooted by using the fungus sequence of *Coprinopsis cinerea* galectin-1 (AF130360.1) as the outgroup. The proteins indicated with asterisk were used for the inference of the previous ancestor (Con-anc). (B) Amino acid sequences of ancestral congerins, ConI, and ConII. Amino acids identical to that of the corresponding last ancestor are represented by a dot.

Figure 1. Amino Acid Sequences and Structures of Ancestral Congerins

http://dx.doi.org/10.5772/56255

This is a protein engineering approach where a proportion of amino acid residues of an extant protein are substituted with those of an ancestor, to construct pseudo-ancestors, in order to reveal the residues determining functional differences between extant and ancestral proteins. These molecular evolutional approaches using pseudo-ancestors bridged from Con-anc to ConI successfully elucidated the regions of the protein relevant to the two adaptive features of ConI, thermostability and higher carbohydrate-binding activity [49]. Experimental molecular archeology analysis, using the reconstructed ancestral congerins, also revealed the process of ConII evolution, another extant galectin. ConII has evolved to enhance affinity for α2, 3-sialyl galactose, which is specifically present in pathogenic marine bacteria. The selection pressure to which Con II reacted was hypothesized to be a shift in carbohydrate affinity. The observed difference in α2, 3-sialyl galactose affinities between Con-anc and Con II support this hypothesis.

The crystal structures of ancestral full-length proteins, Con-anc', Con-anc'-N28K and Con-anc, have been solved at 1.5, 1.6, and 2.0 Å resolutions, respectively [51]. Their three-dimensional (3D) structures clearly demonstrate that Con-anc' or Con-anc'-N28K are appropriate ancestors of extant congerins (Fig. 2). A notable difference between the structures of ConI and ConII is the swapping of S1 strands at the dimer interface, which is unique to ConI among known galectins, and should contribute to its higher stability [81]. The dimer interface of ancestral Con-anc' and Con-anc'-N28K resembled that of ConII, but lacking the strand-swap. This protein-fold is the prototype for dimeric galectins, and the congerin ancestor is expected to have ConII-like conformation. Conversely, Con-anc did display a strand-swapped structure, indicating it was more likely to be an intermediate from the ancestor to ConI, consistent with the results of phylogeny construction (Fig. 2). The differences in carbohydrate interactions between Con-anc' and the extant congerins were observed mainly at the A-face of galactose [51]. These modifications might be relevant to the observed differentiation of carbohydrate specificities between ConI and ConII; ConI prefers α1,4-fucosylated *N*-acetyl glucosamine, while ConII is adapted to bind α2,3-sialyl galactose-containing carbohydrates [49, 50]. Furthermore, structural or functional parameters, such as cytotoxic activity, thermostability of hemagglutination activity, urea and heat denaturation of the structures, and carbohydrate binding activities of the ancestral and extant congerins, were compared as a function of the evolutionary distances from Con-anc' or ConI [51]. Some of these molecular properties were found to be enhanced in both lineages of congerin, which was observed as a correlation with the evolutionary distance from Con-anc'. The dimer interface essential for these proteins to evoke divalent cross-linking activity was enhanced in both lineages as the number of interface H-bonds and dimer interface area increased in ConI and ConII. However, the lactose interface area and the number of lactose H-bonds showed a low correlation with *K*a for carbohydrates, implying that simply enhancing carbohydrate interaction was not likely to be a major selection pressure, and obtaining specificity was more significant for the function of congerins.

Taken together, the first full-length ancestral structures of congerin revealed that the duplicated genes have been differentiating under natural selection pressures for strengthening of the dimer structure and enhancement of the cytotoxic activity. However, the two genes did not react equally to selection pressure, with ConI reacting through protein-fold evolution to

enhance its stability. The modification of the dimer interface in the ConII lineage was rather moderate.

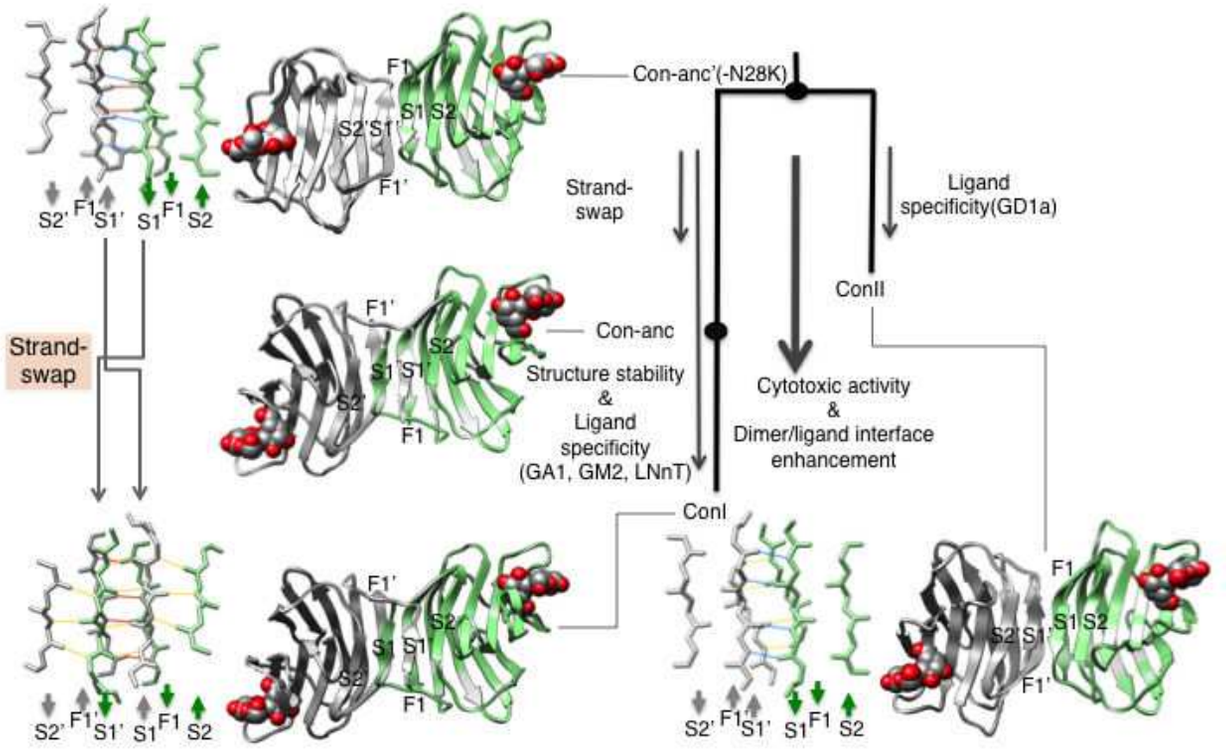


Figure 2. Structures of ConI, ConII, Con-anc and Con-anc'. Con-anc', Con-anc, ConII, and ConI dimers are shown from top to bottom along their molecular phylogeny. The numbers on each branch are the numbers of substitutions. The β strands relevant to the strand-swap at the dimer interface are labeled for S1-S2, and S1′-S2′. Each protein is associated with a close-up of its dimer interface.

5. Reconstruction of ancestral proteins: thermal adaptation of proteins in thermophilic bacterium

Ancestral mutant analysis has been performed to explore the thermal adaptation of proteins. Benner and colleagues reconstructed the ancestral elongation factor-Tu (EF-Tu) predicted using ML methodology, in order to infer the physical environment surrounding ancient organisms [55]. Because EFs play a crucial role in protein synthesis in cells, the thermostability of EFs shows a strong correlation with the optimal growth temperature of their host organisms. For example, the melting temperatures (T_m) of EFs from *Escherichia coli* and *Thermus thermophilus* (HB8) are 42.8°C and 76.7°C, respectively, and the optimal growth temperatures of their

http://dx.doi.org/10.5772/56255

respective hosts are approximately 40°C and 74°C, respectively [86]. Thus, EFs are suitable for use in assessment of the ambient temperature at the time of ancient life. To predict the ancestral sequences of EFs, amino acid sequences of fifty EF-Tu proteins from various bacterial lineages were used to construct two kinds of molecular phylogenetic trees; one using the evolutionary distances calculated using the EF-Tu sequences and the second from distances calculated using ribosomal RNA sequences [87]. Both resulting ancestors had temperature profiles similar to that of the thermophilic EF of modern *Thermus aquaticus*, supporting the hypothesis that the common ancestor of all organisms is a hyperthermophile. Inclusion of additional microbial species into the analysis, and reconstruction of the ancestral EFs at various depths (evolutionary distance from present time) in the phylogeny using the ML method [56, 57], demonstrated that ancestral EFs positioned closer to the root of the phylogenetic tree tended to have significantly higher thermostabilities.

Yamagishi and coworkers reported several ancestral proteins, including two metabolic enzymes; 3-isopropylmalate and dehydrogenase (IPMDH), which is involved in leucine biosynthesis, and isocitrate dehydrogenase (ICDH) involved in the TCA cycle. Ancestral amino acids were introduced into extant IPMDH sequences of the hyperthermophilic archaeon *Sulfolobus tokodaii,* the extremely thermophilic bacterium *Thermus thermophilus*, and the hyperthermophilic archaeon *Caldococcus noboribetus* [14-18].

More recently, Hobbs et al reported the reconstruction of several common Precambrian ancestors of the core metabolic enzyme LeuB, 3-isopropylmalate dehydrogenase, estimated from various *Bacillus* species, in addition to the 3D structure of the last common ancestor at 2.9 Å resolution [19]. Their data indicated that the last common ancestor of LeuB was thermophilic, suggesting that the origin of thermophily in the *Bucillus* genus was ancient. Evolutionary tracing analysis through the ancestors of LeuB also indicated that thermophily was not exclusively a primitive trait, and it could be readily gained as well as lost in evolutionary history [19].

Overall, these studies demonstrate that ancestral enzymes retained enzymatic activity and acquired enhanced thermostability over respective extant enzymes, and that introduction of ancestral state amino acids into modern proteins frequently thermostabilizes them. This indicates that ancestral protein reconstruction can provide empirical access to the evolution of ancient phenotypes, and is useful as a strategy for thermostabilization protein engineering.

6. Reconstruction of ancestral proteins: Evolutionary history of nuclear receptors and visual pigment proteins

Thornton and colleagues have reported seminal work using ancestral protein reconstructions of the nuclear receptors for steroid hormones to investigate evolution of their ligand specificities [38-47, 88, 89]. Vertebrates have six homologous nuclear receptors for steroid hormones; the estrogen receptors alpha and beta (ERα and ERβ), androgen receptor (AR), progesterone receptor (PR), glucocorticoid receptor (GR), and mineralocorticoid receptor (MR). As it is thought that these proteins evolved from a common ancestor through a series of gene dupli-

cations [65], the reconstruction of their ancestral proteins is a useful tool for investigation of their evolution of ligand-specificity. Although GR and MR are close relatives, GR is activated only by the stress hormone cortisol in most vertebrates, while MR is activated by both aldosterone and cortisol [90, 91]. The amino acid substitutions responsible for the specificity of GR toward cortisol were identified by reconstruction studies of the common ancestor of GRs and MRs using ML methodology [38-47]. Thornton and colleagues also reconstructed the ancestral corticoid receptor (AncCR), which corresponded to the protein predicted to be formed at the duplication event between GR and MR genes. Functional analysis showed that AncCR could be activated by both aldosterone and cortisol, suggesting that GR of vertebrates had lost aldosterone specificity during the evolutionary process. Furthermore, site-direct mutagenesis and X-ray crystallographic studies of AncCR revealed that amino acid substitutions at S106P and L111Q were key for the specificity shift of GR [38, 39]. AncCR is the first complete domain ancestor (ligand-binding domain only), for which 3D structure was determined. Ancestral mutant analysis of the NR5 nuclear orphan receptors, including steroidogenic factor 1 (SF-1) and liver receptor homolog 1 (LRH-1) was also reported [41]. The structure-function relationships of the SF-1/LRH-1 subfamily and their evolutionary ligand-binding shift, where the characteristic phospholipid binding ability of the SF-1/LRH-1 subfamily was subsequently reduced and lost in the lineage leading to the rodent LRH-1, due to specific amino acid replacements, were elucidated [41].

Reconstruction of visual pigment proteins, including rhodopsin and green fluorescent protein (GFP)-like proteins, has been also conducted. Chang et al reconstructed an ancestral archosaur rhodopsin from thirty vertebrate species using the ML method and three different models; nucleotide-, amino acid-, and codon-based [25]. An ancestral protein can be reconstructed with each of these models and the inferred archosaur rhodopsin had the same amino acid sequences for all three, except for three amino acid sites (positions 213, 217, and 218), and all reconstructed ancestral proteins had four variants at the ambiguous sites (single mutants T213I, T217A, V218I, and the triple mutant of these) showed similar optical properties, with an apparent absorption maximum at 508–509 nm, slightly red-shifted from that of modern vertebrates (482–507 nm). These data indicated that the alternative ancestral amino acids predicted by the different likelihood models showed similar functional characteristics. Dim-light and color vision in vertebrates are controlled by five visual pigments (RH1, RH2, SWS1, SWS2, and M/LWS), each consisting of a protein moiety (opsin) and a covalently bound 11-cis-retinal (or 11-*cis*-3, 4-dehydroretinal) [91], with characteristic sensitivity to specific wavelengths of maximal absorptions (λ_{max}) from 360 nm (UV) to 560 nm (red). How do the visual pigments achieve sensitivity to various wavelengths? Despite extensive mutagenesis analyses of visual pigments, the molecular mechanisms that modulate the variable λ_{max} values observed in nature were not well understood until ancestral protein reconstruction analysis was applied to the question [92]. Yokoyama and colleagues successfully identified the molecular mechanism of the spectral tuning of visual pigments by generating 15 currently known pigment types using engineered ancestral pigments of SWS1, RH1, and red- and green-sensitive (M/LWS) pigments [26-28]. Kawamura and colleagues reported the reconstruction of ancestral mutants of four green visual pigments from zebrafish, namely, RH2-1, RH2-2, RH2-3 and RH2-4, with λ_{max} values of 467, 476, 488, and 505 nm, respectively [29]. The ancestral pigments showed that

http://dx.doi.org/10.5772/56255

spectral shifts occurred toward the shorter wavelength in evolutionary lineages [29]. Furthermore, Yokoyama and colleagues demonstrated the molecular basis (structural elements) of the adaptation of rhodopsin for the dim-light (deep-sea) vision by ancestral reconstruction experiments using 11 ancestral pigments estimated by rhodopsin sequences of migratory fish from both the surface and deep ocean [30-32].

The great star coral *Montastrea cavernosa* has several green fluorescent protein (GFP)-like proteins, classified into four paralogous groups based on their emission spectra: cyan (emission maximum, 480–495 nm), short wavelength green (500–510 nm), long wavelength green (515–525 nm), and red (575–585 nm) [93]. Matz and colleagues reconstructed the ancestral fluorescent proteins corresponding to the root of each color group, and the common ancestor of the groups using the ML method [33-37]. The analyses of the fluorescence spectra using the ancestral proteins depicted the evolutionary process of the coral GFP-like proteins, in which the peak wavelength has shifted from green to red. Furthermore, they identified the amino acid substitutions responsible for the generation of recent cyan and red fluorescence proteins through site-direct mutagenesis studies of the ancestral green fluorescent protein as a template [35]. Thus, the engineering of ancestral molecules at various evolutionary stages, to recapitulate the changes in their phenotypes over time, is an effective way to explore the molecular evolution and adaptation mechanisms of proteins, although the experimental demonstration of adaptive events at the molecular level is particularly challenging.

7. Concluding remarks

Experimental molecular archaeology using ASR is a new and potentially useful method not only for the study of molecular evolution, but also as a protein engineering technique. This method can provide us with experimental information about ancient genes and proteins, which cannot be obtained from fossil records or by simply constructing molecular phylogeny. However, as discussed above, ancestral sequences can have some issues with ambiguity, depending on the choice of evaluation method, evolutionary model, and sequences. Although inference methods such as MP, ML and BI can lead to errors in predicted ancestral sequences, resulting in potentially misleading estimates of the properties of the ancestral protein, experimental molecular archaeology using ASR could be a more reliable method as all possible ancestral mutants, in which ambiguous amino acid sites are replaced by equally probable candidates individually or in combination, are reproducible and the biological and physicochemical properties and 3D structures of the molecules can be assessed. Indeed, when ancestral congerins were reconstructed based on insufficient sequence information lacking recently determined fish galectin genes, the ancestral Con-anc protein was shown to have a strand-swapped structure resembling ConI, indicating that Con-anc was more likely to be an intermediate mutant of the ancestor to ConI, and that the revised Con-anc' or Con-anc'-N28K are more appropriate ancestors. Thus, the accuracy of ASR can be assessed by analysis of protein activities, stabilities, specificities, and even 3D structures in the laboratory using biochemical or biophysical methods.

Experimental molecular archaeology using ASR can be applied to more complex biological systems, such as heterologous subunit interactions and their evolution in molecular machines [48], host-viral interactions and their co-evolution [54, 94, 95], and proteome/structural proteome level analyses [96]. Furthermore, recent studies have indicated that ASR is applicable to not only to proteins, but also to nucleotides including ancestral rRNA [97] and transposons [95]. To understand the molecular strategies of evolution in nature and the structure-function relationships of proteins and nucleotides, it is important to learn more from 'nature' itself, and from its prodigious works and histories; proteins/nucleotides and their molecular evolution.

Author details

Tomohisa Ogawa[1] and Tsuyoshi Shirai[2]

1 Department of Biomolecular Science, Graduate School of Life Sciences, Tohoku University, Sendai, Japan

2 Nagahama Institute of Bio-Science and Technology, and Japan Science and Technology Agency, Bioinfomatic Research Division, Nagahama, Shiga, Japan

References

[1] Callaway, E. (2010). Fossil genome reveals ancestral link. Nature 468 (7327), 1012.

[2] Pääbo, S, Poinar, H, Serre, D, Jaenicke-després, V, Hebler, J, Rohland, N, Kuch, M, Krause, J, Vigilant, L, & Hofreiter, M. (2004). Genetic analyses from ancient DNA. Ann. Rev. Genet. , 38, 645-679.

[3] Green, R. E, Krause, J, Briggs, A. W, Maricic, T, Stenzel, U, Kircher, M, Patterson, N, Li, H, Zhai, W, Fritz, M. H-Y, Hansen, N. F, Durand, E. Y, Malaspinas, A. S, Jensen, J. D, Marques-bonet, T, Alkan, C, Prüfer, K, Meyer, M, Burbano, H. A, Good, J. M, Schultz, R, Aximu-petri, A, Butthof, A, Höber, B, Höffner, B, Siegemund, M, Weihmann, A, Nusbaum, C, Lander, E. S, Russ, C, Novod, N, Affourtit, J, Egholm, M, Verna, C, Rudan, P, Brajkovic, D, Kucan, Ž, Gušic, I, Doronichev, V. B, Golovanova, L. V, Lalueza-fox, C, De La Rasilla, M, Fortea, J, Rosas, A, & Schmitz, R. W. Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. ((2010). A draft sequence of the neandertal genome. Science 328 (5979), 7710-7722.

[4] Krause, J, Fu, Q, Good, J. M, Viola, B, Shunkov, M. V, Derevianko, A. P, & Pääbo, S. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. ((2010). Nature, 464 (7290), 894-897.

[5] Reich, D, Green, R. E, Kircher, M, Krause, J, Patterson, N, Durand, E. Y, Viola, B, Briggs, A. W, & Stenzel, U. Johnson PLF, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin J-J, Kelso J, Slatkin M, Pääbo S. ((2010). Genetic history of an archaic hominin group from Denisova cave in Siberia. Nature 468 (7327), 1053-1060.

[6] Rohland, N, Reich, D, Mallick, S, Meyer, M, Green, R. E, Georgiadis, N. J, Roca, A. L, & Hofreiter, M. (2010). Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. PLoS biol., 8 (12), e1000564

[7] Gibson, D. G, Glass, J. I, Lartigue, C, Noskov, V. N, Chuang, R-Y, Algire, M. A, Benders, G. A, Montague, M. G, Ma, L, Moodie, M. M, Merryman, C, Vashee, S, Krishnakumar, R, Assad-garcia, N, Andrews-pfannkoch, C, Denisova, E. A, Young, L, Qi, Z-Q, Segall-shapiro, T. H, Calvey, C. H, & Parmar, P. P. Hutchison III CA, Smith HO, Venter JC. ((2011). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome Science 329 (5987), 52-56.

[8] Lai, J, Jin, J, Kubelka, J, & Liberles, D. A. (2012). A Phylogenetic Analysis of Normal Modes Evolution in Enzymes and Its Relationship to Enzyme Function. J. Mol. Biol. , 422, 442-459.

[9] Harms, M. J, & Thornton, J. W. (2010). Analyzing protein structure and function using ancestral gene reconstruction. Curr. Opin. Struct. Biol. 20 (3), 360-366.

[10] Stackhouse, J, Presnell, S. R, Mcgeehan, G. M, Nambiar, K. P, & Benner, S. A. (1990). The ribonuclease from an extinct bovid ruminant. FEBS Lett., , 262, 104-106.

[11] Jermann, T. M, Opitz, J. G, Stackhouse, J, & Benner, S. A. (1995). Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. Nature , 374, 57-59.

[12] Malcolm, B. A, Wilson, K. P, Matthews, B. W, Kirsch, J. F, & Wilson, A. C. (1990). Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. Nature , 345, 86-89.

[13] Thomson, J. M, Gaucher, E. A, Burgan, M. F, De Kee, D. W, Li, T, Aris, J. P, & Benner, S. A. (2005). Resurrecting ancestral alcohol dehydrogenases from yeast. Nat. Genet. , 37, 630-635.

[14] Miyazaki, J, Nakaya, S, Suzuki, T, Tamakoshi, M, Oshima, T, & Yamagishi, A. (2001). Ancestral residues stabilizing isopropylmalate dehydrogenase of an extreme thermophile: Experimental evidence supporting the thermophilic common ancestor hypothesis. J. Biochem. 129 (5), 777-782., 3.

[15] Watanabe, K, Ohkuri, T, Yokobori, S, & Yamagishi, A. (2006). Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. J. Mol. Biol. , 355, 664-674.

[16] Watanabe, K, & Yamagishi, A. (2006). The effects of multiple ancestral residues on the Thermus thermophilus isopropylmalate dehydrogenase. FEBS Lett. 580 (16), 3867-3871., 3.

[17] Dean, A. M, & Golding, G. B. (1997). Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. Proc. Natl. Acad. Sci. USA 94 (7), 3104-3109.

[18] Iwabata, H, Watanabe, K, Ohkuri, T, Yokobori, S, & Yamagishi, A. (2005). Thermostability of ancestral mutants of Caldococcus noboribetus isocitrate dehydrogenase. FEMS Microbiol. Lett., , 243, 393-398.

[19] Hobbs, J. K, Shepherd, C, Saul, D. J, Demetras, N. J, Haaning, S, Monk, C. R, Daniel, R. M, & Arcus, V. L. (2012). On the origin and evolution of thermophily: Reconstruction of functional precambrian enzymes from ancestors of Bacillus. Mol. Biol. Evol. 29 (2), 825-835.

[20] Chandrasekharan, U. M, Sanker, S, Glynias, M. J, Karnik, S. S, & Husain, A. (1996). Angiotensin II-forming activity in a reconstructed ancestral chymase. Science , 271, 502-505.

[21] Akanuma, S, Iwami, S, Yokoi, T, Nakamura, N, Watanabe, H, Yokobori, S-I, & Yamagishi, A. (2011). Phylogeny-based design of a B-subunit of DNA gyrase and its ATPase domain using a small set of homologous amino acid sequences. J. Mol. Biol. 412 (2), 212-225.

[22] Shimizu, H, Yokobori, S-i, Ohkuri, T, Yokogawa, T, Nishikawa, K, & Yamagishi, A. (2007). Extremely Thermophilic Translation System in the Common Ancestor Commonote: Ancestral Mutants of Glycyl-tRNA Synthetase from the Extreme Thermophile Thermus thermophiles. J. Mol. Biol. 369 (4), 1060-1069.

[23] Alcolombri, U, Elias, M, & Tawfik, D. S. (2011). Directed evolution of sulfotransferases and paraoxonases by ancestral libraries. J. Mol. Biol. 411 (4), 837-853.

[24] Perez-jimenez, R, Inglés-prieto, A, Zhao, Z-M, Sanchez-romero, I, Alegre-cebollada, J, Kosuri, P, Garcia-manyes, S, Kappock, T. J, Tanokura, M, Holmgren, A, Sanchez-ruiz, J. M, Gaucher, E. A, & Fernandez, J. M. (2011). Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. Nature Struct. Mol. Biol. 18 (5), 592-596.

[25] Chang, B. S, Jonsson, K, Kazmi, M. A, Donoghue, M. J, & Sakmar, T. P. (2002). Recreating a functional ancestral archosaur visual pigment. Mol. Biol. Evol. , 19, 1483-1489.

[26] Shi, Y, & Yokoyama, S. (2003). Molecular analysis of the evolutionary significance of ultraviolet vision in vertebrates. Proc. Natl. Acad. Sci. USA. , 100, 8308-8313.

[27] Yokoyama, S, & Takenaka, N. (2004). The molecular basis of adaptive evolution of squirrelfish rhodopsins. Mol. Biol. Evol. , 21, 2071-2078.

[28] Yokoyama, S, Yang, H, & Starmer, W. T. (2008). Molecular basis of spectral tuning in the red- and green-sensitive (M/LWS) pigments in vertebrates. Genetics , 179, 2037-2043.

[29] Chinen, A, Matsumoto, Y, & Kawamura, S. (2005). Reconstitution of ancestral green visual pigments of zebrafish and molecular mechanism of their spectral differentiation. Mol. Biol. Evol., , 22, 1001-1010.

[30] Yokoyama, S, Tada, T, Zhang, H, & Britt, L. (2008). Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. Proc. Natl. Acad. Sci. USA., , 105, 13480-13485.

[31] Yokoyama, S. (2008). Evolution of dim-light and color vision pigments. Annu. Rev. Genom. Hum Genet., , 9, 259-282.

[32] Watanabe, H. C, Mori, Y, Tada, T, Yokoyama, S, & Yamato, T. (2010). Molecular mechanism of long-range synergetic color tuning between multiple amino acid residues in conger rhodopsin. Biophysics , 6, 67-78.

[33] Ugalde, J. A, Chang, B. S, & Matz, M. V. (2004). Evolution of coral pigments recreated. Science, 305, 1433.

[34] Chang BSWUgalde JA, Matz MV. ((2005). Applications of ancestral protein reconstruction in understanding protein function: GFP-like proteins. Method Enzymol. , 395, 652-670.

[35] Field, S. F, Bulina, M. Y, Kelmanson, I. V, Bielawski, J. P, & Matz, M. V. (2006). Adaptive evolution of multicolored fluorescent proteins in reef-building corals. J. Mol. Evol. 62 (3), 332-339.

[36] Alieva, N. O, Konzen, K. A, Field, S. F, Meleshkevitch, E. A, Hunt, M. E, Beltran-ramirez, V, Miller, D. J, Wiedenmann, J, Salih, A, & Matz, M. V. (2008). Diversity and evolution of coral fluorescent proteins. PLoS ONE 3 (7), art. (e2680)

[37] Field, S. F, & Matz, M. V. (2010). Retracing evolution of red fluorescence in GFP-like proteins from faviina corals. Mol. Biol. Evol. 27 (2), 225-233.

[38] Thornton, J. W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc. Natl. Acad. Sci. USA, , 98, 5671-5676.

[39] Thornton, J. W, Need, E, & Crews, D. (2003). Resurrecting the ancestral steroid receptor: Ancient origin of estrogen signaling. Science 301 (5640), 1714-1717.

[40] Thornton, J. W. (2004). Resurrecting ancient genes: Experimental analysis of extinct molecules. Nat. Rev. Genet. 5 (5), 366-375.

[41] Krylova, I. N, Sablin, E. P, Moore, J, Xu, R. X, & Waitt, G. M. MacKay JA, Juzumiene D, Bynum JM, Madauss K, Montana V, Lebedeva L, Suzawa M, Williams JD, Williams SP, Guy RK, Thornton JW, Fletterick RJ, Willson TM, Ingraham HA. ((2005).

Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. Cell 120 (3), 343-355.

[42] Bridgham, J. T, Carroll, S. M, & Thornton, J. W. (2006). Evolution of hormone-receptor complexity by molecular exploitation. Science 312 (5770), 97-101.

[43] Ortlund, E. A, Bridgham, J. T, Redinbo, M. R, & Thornton, J. W. (2007). Crystal structure of an ancient protein: Evolution by conformational epistasis. Science 317 (5844), 1544-1548.

[44] Bridgham, J. T, Ortlund, E. A, & Thornton, J. W. (2009). An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461 (7263), 515-519

[45] Bridgham, J. T, Eick, G. N, Larroux, C, Deshpande, K, & Harms, M. J. Gauthier MEA, Ortlund EA, Degnan BM, Thornton JW. ((2010). Protein evolution by molecular tinkering: Diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. PLoS Biology 8 (10), art. (e1000497)

[46] Eick, G. N, & Thornton, J. W. (2011). Evolution of steroid receptors from an estrogen-sensitive ancestral receptor. Mol. Cell. Endocrinol. 334 (1-2), 31-38.

[47] Carroll, S. M, Ortlund, E. A, & Thornton, J. W. (2011). Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. PLoS Genetics 7 (6), art. (e1002117)

[48] Finnigan, G. C, Hanson-smith, V, Stevens, T. H, & Thornton, J. W. (2012). Evolution of increased complexity in a molecular machine. Nature 481 (7381), 360-364.

[49] Konno, A, Ogawa, T, Shirai, T, & Muramoto, K. (2007). Reconstruction of a probable ancestral form of conger eel galectins revealed their rapid adaptive evolution process for specific carbohydrate recognition. Mol. Biol. Evol., , 24, 2504-2514.

[50] Konno, A, Yonemaru, S, Kitagawa, A, Muramoto, K, Shirai, T, & Ogawa, T. (2010). Protein engineering of conger eel galectins by tracing of molecular evolution using probable ancestral mutants. BMC Evol. Biol., 10:43, doi:10.1186/1471-2148-10-43.

[51] Konno, A, Kitagawa, A, Watanabe, M, Ogawa, T, & Shirai, T. (2011). Tracing protein evolution through ancestral structures of fish galectin.
Structure 19 (5), 711-721.

[52] Yadid, I, & Tawfik, D. S. (2011). Functional-propeller lectins by tandem duplications of repetitive units. Protein Eng. Design Select. 24 (1-2), 185-195.

[53] Gullberg, M, Tolf, C, Jonsson, N, Mulders, M. N, Savolainen-kopra, C, Hovi, T, Van Ranst, M, Lemey, P, Hafenstein, S, & Lindberg, A. M. Characterization of a putative ancestor of coxsackievirus B5. J. Virology 84 (19), 9695-9708.

[54] Kaiser, S. M, Malik, H. S, & Emerman, M. (2007). Restriction of an extinct retrovirus by the human TRIM5α antiviral protein. Science , 316, 1756-1758.

[55] Gaucher, E. A, Thomson, J. M, Burgan, M. F, & Benner, S. A. (2003). Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature 425 (6955), 285-288.

[56] Gaucher, E. A, Govindarajan, S, & Ganesh, O. K. (2008). Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature 451 (7179), 704-707.

[57] Gouy, M, & Chaussidon, M. (2008). Evolutionary biology: ancient bacteria liked it hot. Nature, , 451, 635-636.

[58] Whittington, A. C, & Moerland, T. S. (2012). Resurrecting prehistoric parvalbumins to explore the evolution of thermal compensation in extant Antarctic fish parvalbumins. J. Exp. Biol. 215 (18), 3281-3292.

[59] Hult, E. F, Weadick, C. J, Chang, B. S. W, & Tobe, S. S. (2008). Reconstruction of ancestral FGLamide-type insect allatostatins: A novel approach to the study of allatostatin function and evolution. J. Insect Physiol. 54 (6), 959-968.

[60] Skovgaard, M, Kodra, J. T, Gram, D. X, Knudsen, S. M, Madsen, D, & Liberles, D. A. (2008). Using Evolutionary Information and Ancestral Sequences to Understand the Sequence-Function Relationship in GLP-1 Agonists J.Mol. Biol. 363 (5), 977-988.

[61] Beintema, J. J. (1987). Structure, properties and molecular evolution of pancreatic-type ribonucleases. Life Chem. Rep., , 4, 333-389.

[62] Carroll, R. L. (1988). Vertebrate Paleontology and Evolution, New York, Freeman.

[63] Woese, C. R. (1987). Bacterial evolution. Microbiol. Rev., , 51, 221-271.

[64] Imada, K, Sato, M, Tanaka, N, Katsube, Y, Matsuura, Y, & Oshima, T. (1991). Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of Thermus thermophilus at 2.2 A resolution. J. Mol. Biol., , 222, 725-738.

[65] Eck, R. V, & Dayhoff, M. O. (1966). Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. Science, , 152, 363-366.

[66] Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol., , 17, 368-376.

[67] Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol., , 24, 1586-1591.

[68] Pupko, T. Pe'er, I., Shamir, R. & Graur, D. ((2000). A fast algorithm for joint reconstruction of ancestral amino acid sequences. Mol. Biol. Evol. , 17, 890-896.

[69] Cai, W, Pei, J, & Grishin, N. V. (2004). Reconstruction of ancestral protein sequences and its applications. BMC Evol. Biol., 4, 33.

[70] Edwards, R. J, & Shields, D. C. (2004). GASP: Gapped ancestral sequence prediction for proteins. BMC Bioinf., 5, 123.

[71] Williams, P. D, Pollock, D. D, Blackburne, B. P, & Goldstein, R. A. (2006). Assessing the accuracy of ancestral protein reconstruction methods. PLoS Comput. Biol. 2 (6), 0598-0605.

[72] Hanson-smith, V, Kolaczkowski, B, & Thornton, J. W. (2010). Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. Mol. Biol. Evol. 27 (9), 1988-1999.

[73] Kaminuma, E, Mashima, J, Kodama, Y, Gojobori, T, Ogasawara, O, Okubo, K, Takagi, T, & Nakamura, Y. (2010). DDBJ launches a new archive database with analytical tools for next-generation sequence data. Nucleic Acids Res. 38, DD38., 33.

[74] Katoh, K, Misawa, K, Kuma, K, & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. , 30, 3059-3066.

[75] Yang, Z, Nielsen, R, Goldman, N, & Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics , 155, 431-449.

[76] Kamiya, H, Muramoto, K, & Goto, R. (1988). Purification and properties of agglutinins from conger eel, Conger myriaster (Brevoort), skin mucus. Dev. Comp. Immuno., , 12, 309-318.

[77] Muramoto, K, & Kamiya, H. (1992). The amino-acid sequence of a lectin from conger eel, Conger myriaster, skin mucus. Biochem. Biophys. Acta, , 1116, 129-136.

[78] Muramoto, K, Kagawa, D, Sato, T, Ogawa, T, Nishida, Y, & Kamiya, H. (1999). Functional and structural characterization of multiple galectins from the skin mucus of conger eel, Conger myriaster. Comp. Biochem. Physiol. B, , 123, 33-45.

[79] Nakamura, O, Matsuoka, H, Ogawa, T, Muramoto, K, Kamiya, H, & Watanabe, T. (2006). Opsonic effect of congerin, a mucosal galectin of the Japanese conger, Conger myriaster (Brevoort). Fish Shellfish Immunol. , 20, 433-435.

[80] Ogawa, T, Ishii, C, Kagawa, D, Muramoto, K, & Kamiya, H. (1999). Accelerated evolution in the protein-coding region of galectin cDNAs, congerin I and congerin II, from skin mucus of conger eel (Conger myriaster). Biosci. Biotechnol. Biochem., , 63, 1203-1208.

[81] Ogawa, T, Shirai, T, Shionyu-mitsuyama, C, Yamane, T, Kamiya, H, & Muramoto, K. (2004). The speciation of conger eel galectins by rapid adaptive evolution. Glycoconj. J., , 19, 451-458.

[82] Shirai, T, Mitsuyama, C, Niwa, Y, Matsui, Y, Hotta, H, Yamane, T, Kamiya, H, Ishii, C, Ogawa, T, & Muramoto, K. (1999). High-resolution structure of the conger eel galectin, congerin I, in lactose-liganded and ligand-free forms: emergence of a new structure class by accelerated evolution. Structure, , 7, 1223-1233.

[83] Shirai, T, Matsui, Y, Shionyu-mitsuyama, C, Yamane, T, Kamiya, H, Ishii, C, Ogawa, T, & Muramoto, K. (2002). Crystal structure of a conger eel galectin (Congerin II) at

1.45 angstrom resolution: Implication for the accelerated evolution of a new ligand-binding site following gene duplication. J. Mol. Biol., , 321, 879-889.

[84] Shirai, T, Shionyu-mitsuyama, C, Ogawa, T, & Muramoto, K. (2006). Structure based studies of the adaptive diversification process of congerins. Mol. Div., , 10, 567-573.

[85] Ogawa, T. (2006). Molecular diversity of proteins in biological offense and defense systems. Mol. Div., , 10, 511-514.

[86] Williams, R. A. D. da Costa, M. S. ((1992). The genus Thermus and related microorganisms. In Balows, A. Truper, H. G., Dworkin, M., Harder, W. & Schleifer, K.-H. (Eds.), The Prokaryotes (2nd edition, New York, Springer., 3745-3753.

[87] Hugenholtz, P, Goebel, B. M, & Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol., , 180, 4765-4774.

[88] Thornton, J. W, & Desalle, R. (2000). A new method to localize and test the significance of incongruence: detecting domain shuffling in the nuclear receptor superfamily. Syst. Biol., , 49, 183-201.

[89] Dean, A. M, & Thornton, J. W. (2007). Mechanistic approaches to the study of evolution: The functional synthesis. Nat. Rev. Genet. 8(9), 675-688.

[90] Bentley, P. J. (1998). Comparative Vertebrate Endocrinology, Cambridge Univ. Press, Cambridge.

[91] Farman, N, & Rafestin-oblin, M. E. (2001). Multiple aspects of mineralocorticoid selectivity. Am. J. Physiol. Renal. Physiol., 280. F , 181-192.

[92] Yokoyama, S. (2000). Molecular evolution of vertebrate visual pigments. Prog. Retin. Eye Res. , 19, 385-419.

[93] Kelmanson, I. V, & Matz, M. V. (2003). Molecular basis and evolutionary origins of color diversity in great star coral Montastraea cavernosa (Scleractinia: Faviida). Mol. Biol. Evol., , 20, 1125-1133.

[94] Heinemann, J, Maaty, W. S, Gauss, G. H, Akkaladevi, N, Brumfield, S. K, Rayaprolu, V, Young, M. J, Lawrence, C. M, & Bothner, B. (2011). Fossil record of an archaeal HK97-like provirus. Virology , 417, 362-368.

[95] Münk, C, Willemsen, A, & Bravo, I. G. (2012). An ancient history of gene duplications, fusions andlosses in the evolution of APOBEC3 mutators inmammals. BMC Evol. Biol., 12, 71 doi:10.1186/1471-2148-12-71.

[96] Erdin, S, Ward, R. M, Venner, E, & Lichtarge, O. (2010). Evolutionary Trace Annotation of Protein Function in the Structural Proteome. J. Mol. Biol. , 396, 1451-1473.

[97] Lu, Q, & Fox, G. E. Resurrection of an ancestral 5S rRNA ((2011). BMC Evol. Biol. 11 (1), art. (218)

IntechOpen