IV

Experimental methodology and concerns

CHAPTER 14

Experimental resurrection of ancient biomolecules: gene synthesis, heterologous protein expression, and functional assays

Eric A. Gaucher

14.1 Introduction

The recent accumulation of DNA sequence data, combined with advances in evolutionary theory and computational power, have paved the way for innovative approaches to understand the origins, evolution, and distribution of life and its constituent biomolecules. One approach to understanding ancestral states follows a present-day-backwards strategy, whereby genomic sequences from extant (modern) organisms are incorporated into evolutionary models that estimate the extinct (ancient) character states of genes no longer present on Earth. These inferred ancestral gene sequences act as hypotheses that can be tested in the laboratory through the resurrection of the ancestral proteins themselves (Pauling and Zuckerkandl, 1963; Thornton, 2004). Results from functional assays of the protein products from these ancient genes permit us to accept/reject hypotheses about the sequences themselves, or about their interactions, binding specificities, environments, etc.

To date, approximately 20 narratives have emerged where specific molecular systems from extinct organisms have been resurrected for study in the laboratory (Sassi *et al*., 2007). These include digestive proteins (ribonucleases, proteases, and lysozymes) in ruminants and primates to illustrate how digestive function arose from non-digestive function in response to a changing global ecosystem (Jermann *et al*., 1995; Zhang, 2006), fermentive enzymes from fungi to illustrate how molecular adaptation supported mammals as they displaced dinosaurs as the dominant large land animals (Thomson *et al*., 2005), pigments in the visual system adapting to different environments (Chang *et al*., 2002; Shi and Yokoyama, 2003; Chinen *et al*., 2005), steroid hormone receptors adapting to changing function in steroid-based regulation of metazoans (Thornton *et al*., 2003), fluorescent proteins from ocean-dwelling invertebrates (Ugalde *et al*., 2004), enzyme cofactor evolution (Zhu *et al*., 2005), proteins from very ancient bacteria helping to define environments where the earliest forms of bacterial life lived (Miyazaki *et al*., 2001; Gaucher *et al*., 2003; Iwabata *et al*., 2005), among others.

This chapter will summarize the different approaches exploited by these studies. This includes the different strategies exploited when building ancient genes in the laboratory, the various systems used to express the encoded proteins of the ancient genes, and the different types of functional assay used to characterize the behaviors of the ancient biomolecules.

14.2 Constructing ancient sequences in the laboratory

14.2.1 Site-directed mutagenesis

A widely used approach to synthesize ancestral genes is site-directed mutagenesis. This is the

optimal approach when an ancestral gene can be obtained by introducing a small number of mutations into a modern (extant) form of the gene. This approach has been used, for example, in the synthesis of ancestral alcohol dehydrogenase (ADH) and seminal ribonuclease family members discussed in other chapters of this book, as well as other families such as isocitrate dehydrogenase and eosinophil-derived neurotoxins (EDNs; Zhang and Rosenberg, 2002; Zhu *et al*., 2005).

In my experience, the key to a successful mutagenesis experiment lies in the design of primers. For instance, the QuikChange® site-directed mutagenesis protocol (Stratagene) recommends multiple conditions be met when designing primers. The primers should be between 25 and 45 bases in length to avoid unwanted secondary-structural elements and have melting temperatures above 75°C. Furthermore, the primers should have a minimal G + C content of 40%, terminate in one or more Cs and/or Gs, and the site-specific mutations should lie in the middle of the primer.

After proper primer design, mutations are achieved through thermal cycling using template DNA (backbone gene), a polymerase with high fidelity and the primer pairs. The parental (non-mutated) DNA can then be degraded with *Dpn*I due to the methylated/hemimethylated states of the template DNA if isolated from an organismal host. Amplification of the mutated DNA is achieved using standard cloning and transformation protocols.

14.2.2 Gene synthesis

Site-directed mutagenesis is a rapid and cost-effective way to generate mutants so long as the number of required mutations is manageable. When the number of mutations is large, however, this approach holds less value. Under these circumstances, an alternative approach is to synthesize the complete ancestral gene using oligonucleotides and primer-extension reactions. Again, primer design is critical to achieve successful gene synthesis. Two approaches for primer design have been used to synthesize ancestral genes and are discussed below. Both approaches follow the same general protocol but differ in the size of oligonucleotides used to synthesize the ancestral genes (Figure 14.1).

Incorporation of short oligonucleotides: elongation factors

The first example of total synthesis for an enzyme-encoded gene was presented more than 20 years ago. These authors used 66 different oligonucleotides ranging in size from 10 to 22 bases to synthesize a ribonuclease S gene (Nambiar *et al*., 1984). Although advances in molecular biology techniques and reagents have greatly enhanced this approach, the basic principles guiding it are still practical. In short, overlapping primers are used to generate a backbone of the gene, a DNA polymerase uses deoxynucleotide triphosphates (dNTPs) as a substrate to insert bases, thereby

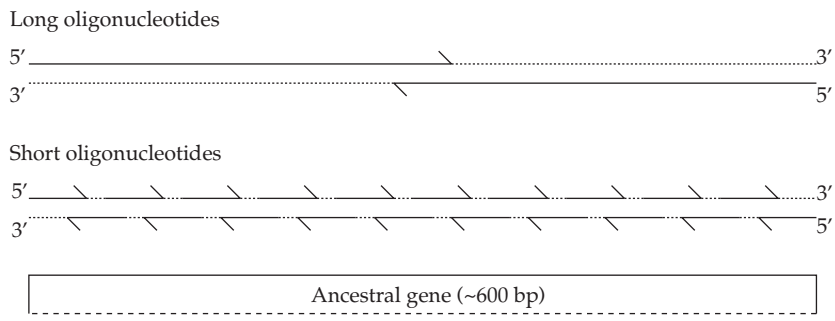

Figure 14.1 Schematic of gene synthesis using long and short oligonucleotides for ancestral genes. Long oligonucleotides are approximately 300 bases in length and short nucleotides are typically 50 bases in length. The primers overlap across 15–20 bases regardless of whether using long or short oligonucleotides. Arrows indicate 5′ → 3′ directionality. To produce full-length genes, or large fragments of the gene, primers are mixed in a PCR reaction and subjected to one extension cycle, thereby allowing the polymerase to synthesize the second strand of the non-base-paired regions of the gene (shown as dashed lines). Amplification of the gene product is achieved using flanking primers and standard PCR conditions and cycles.

filling in the gaps of the backbone, and then PCR is used to amplify the whole gene.

Total synthesis of ancient bacterial elongation factors followed this procedure (see Chapter 2 in this volume). The ancestral genes were synthesized in two steps by thermal reactions using complementary 50-mer oligonucleotides with 15–20-bp overlap. Nineteen primers were used to synthesize the 3′ end of the gene while an additional 18 primers were used to synthesize the 5′ end of the gene. Each reaction consisted of 10 × PCR buffer, 10 mM dNTPs, 5 μM each primer, and 2.5 units of *Pfu* polymerase in a 25-μl reaction (*Pfu* has greater proofreading ability and higher processivity than *Taq* polymerase). The reactions were heated to 94°C for 3 min, cooled to 60°C (1°C/10 s) for 10 min, and heated to 72°C for 5 min. The products of the two reactions then served as template DNA for a PCR reaction. This reaction consisted of 10 × PCR buffer, 10 mM dNTPs, 10 μM each of two flanking primers, 2.5 μl of a 1 : 10 dilution of each of the two template DNA reactions, 2.5 units of IDProof polymerase (ID Labs Biotechnology). This reaction mixture was amplified by PCR using the following settings: 94°C for a 1-min hot-start, 94°C for 30 s, 60°C for 45 s, and 72°C for 90 s, with 10 cycles to minimize errors during the amplification process. Genes were cloned in the Topo-TA vector system (Invitrogen). Errors resulting from primer synthesis and/or PCR were fixed using the standard site-directed mutagenesis protocol discussed above. In my experience, it seems inevitable that mutations are present in the completely synthesized gene and that most of these mutations arise from errors during synthesis of oligonucleotides.

Incorporation of long oligonucleotides: rhodopsins

Longer oligonucleotides offer some advantages over short oligonucleotides when designing ancestral genes. For instance, short oligonucleotides require on the order of 30–100 overlapping fragments (depending on gene size), and it may be difficult to optimize the GC content, melting temperature, etc., for each primer. Since regions of overlap are required to be about 20 bp in length when building genes, regardless of whether employing short or long oligonucleotides, fewer long oligonucleotides are required to cover an entire synthetic gene. Incorporating long oligonucleotides therefore minimizes the total amount of overlap required and reduces the total number of bases to be synthesized. This strategy is potentially more economical, but this remains to be determined. In this regard, the accuracy of oligonucleotide synthesis is substantially diminished when the DNA fragment is longer than approximately 50–60 bases. If a 300-mer oligonucleotide contains numerous incorrect bases, than the overall cost of synthesizing an ancestral gene will be higher than anticipated due to the increased cost of repairing the gene via site-directed mutagenesis.

Regardless of this concern, long oligonucleotides have been successfully employed to synthesize an ancestral rhodopsin gene inferred from vertebrate species (Chang *et al*., 2002). The entire ancestral gene (1114 bp long) was synthesized from five long oligonucleotides and amplified using PCR. A detailed discussion of this approach is presented elsewhere in the book (see Chapter 15).

14.2.3 Codon optimization

It is well known that codon preference differs between species and that there is a correlation between codon usage of highly expressed genes and tRNA molecules within individual species. Although the exact nature of this correlation is unclear, it is important to take it into account when constructing ancient sequences. Expression of an ancient metazoan gene in microorganisms such as *Escherichia coli* can result in misincorporation of certain amino acids during protein translation if the metazoan codons are not optimized for bacterial expression. For instance, the AGG codon for arginine is frequently used in metazoan species such as *Drosophila* and humans. In *E. coli*, however, this codon is used less frequently than the most common stop codon. Incorporating this codon in an ancient gene inferred from metazoan sequences but expressed in *E. coli* may result in misincorporation of amino acids, translational frameshifts, and an overall reduction in translational efficiency.

Interestingly, observing differences in codon usage does not require an inter-domain comparison

of species, such as bacteria and eukaryotes. Even the expression of a bacterial gene in a heterologous bacterial species can be affected by codon usage. Figure 14.2a shows the differences in expression profiles for a *Taq* polymerase (from *Thermus aquaticus*) translated in *E. coli* using wild-type *Thermus* codons compared with a synthesized *Thermus* polymerase codon optimized for expression in *E. coli*. Such results highlight the importance of considering codon

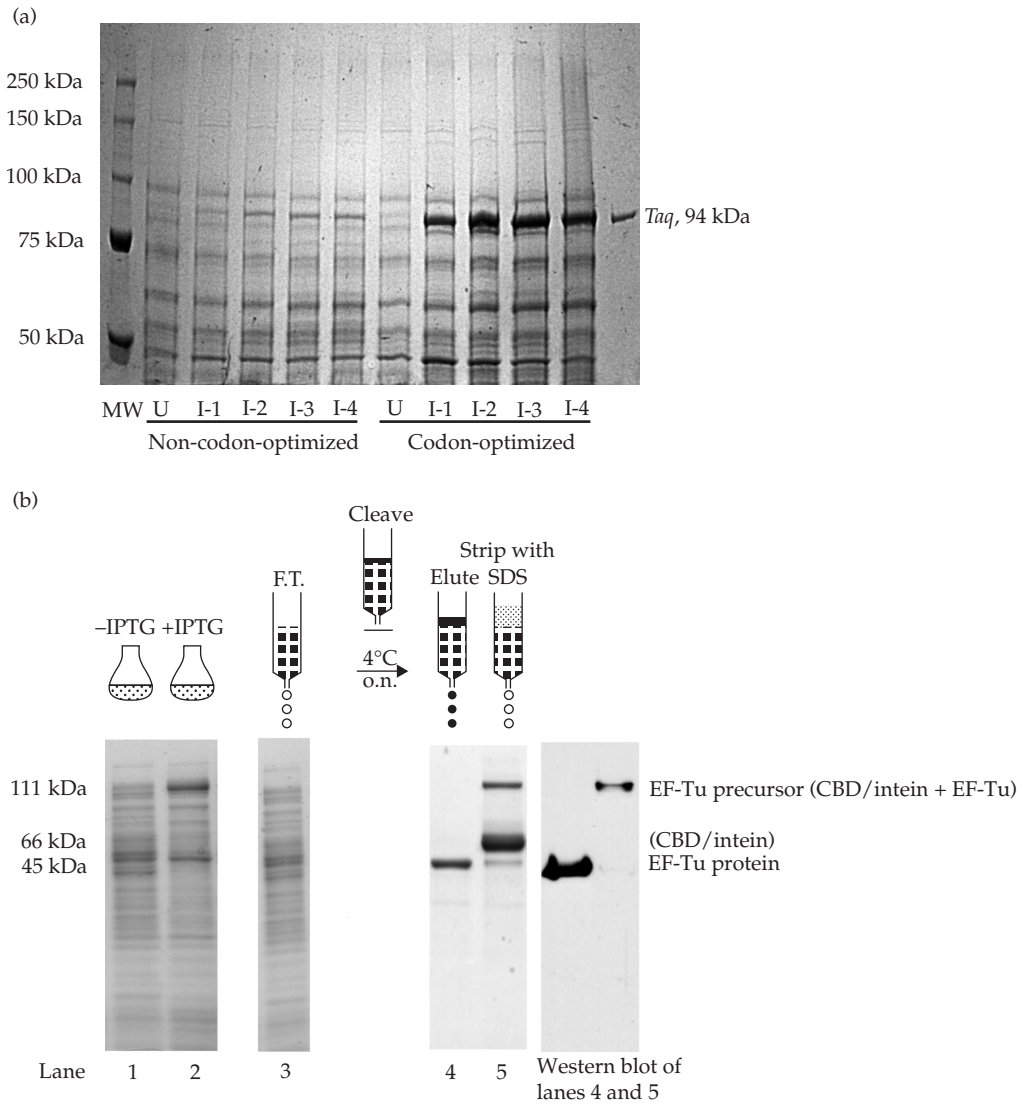

Figure 14.2 Heterologous protein expression and purification. (a) The effects of codon-optimized and non-codon-optimized heterologous gene expression of *Thermus aquaticus* polymerase (*T*aq) in E. coli cells. Molecular weight markers are shown (MW). Lanes include non-induced cells (U) and induced cells after 1, 2, 3, and 4 h of expression (I-1–I-4). (b) Procedure for protein purification of modern and ancestral elongation factor (EF) proteins. EF protein is expressed as a construct with a chitin-binding domain (CBD) and an intein domain (a self-cleaving peptide that releases the EF protein). Shown is an example of non-induced cells (lane 1), induced cells (lane 2), flow-through (F.T.) of the column indicating strong binding of the construct to the column (lane 3), release of EF protein upon activation of the intein domain (lane 4), protein present on column after elution (lane 5), and Western-blot hybridization using an antibody that recognizes a conserved epitope in modern and ancestral EF proteins (right-hand column). IPTG, isopropyl β-D-thiogalactoside; o.n., overnight; SDS, sodium dodecyl sulfate.

optimization when expressing ancient genes in heterologous species, as this optimization may be required to obtain sufficient quantities of accurately translated protein for functional assays.

14.3 Heterologous expression

Regardless of whether ancestral genes are synthesized using site-directed mutagenesis, overlapping PCR, or the currently popular outsourcing to a gene-synthesis company (often charging US$0.60–1.30 per synthesized base pair), the majority of resurrection studies are more concerned with the ancient protein than the ancient DNA itself. Upon completion of gene synthesis, ancestral genes are typically translated in an expression host *in vivo*. The use of specific expression hosts is often determined by known protein-folding/-activation requirements (redox potential, molecular chaperones, post-translational modifications, propensity to precipitate, etc.) and whether ancestral protein behaviors are determined through *in vitro* or *in vivo* cellular assays.

14.3.1 Ancestral bacterial gene expressed in bacterial cells

The ancestral elongation factor genes discussed earlier in this chapter were cloned into and expressed from the TYB11 vector in *E. coli* (IMPACT System; intein-mediated purification with an affinity chitin-binding tag), and purified according to the manufacturer's instructions (New England Biolabs). The proteins were eluted from a chitin-affinity column in a buffer consisting of 20 mM Tris/HCl (pH 8.5), 500 mM NaCl, 10 mM $MgCl_2$, 5 μM GDP, and 1 μM phenylmethylsulfonyl fluoride (PMSF), and stored at −20°C. The samples were filtered and concentrated using Centricon YM-30 (Amicon). SDS/PAGE verified the isolation of a single band of appropriate size, approximately 44 kDa (Figure 14.2b). Western blots were performed using a monoclonal antibody (mAb 900) that recognizes a conserved epitope from extant elongation factor (EF) proteins and present in the ancestral proteins.

14.3.2 Ancestral yeast gene expressed in yeast cells

Isogenic strains of *Saccharomyces cerevisiae*, BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) and BY4742 (MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0; A.T. C.C.), were used the create deletions of the two primary alcohol dehydrogenase alleles ADH2 and ADH1, respectively (Thomson *et al.*, 2005). Primers were designed to amplify the URA3 gene with 50 bp of sequence identity to either the ADH1 or ADH2 5′ and 3′ untranslated regions. The URA3-containing vector pRS316 was used as a template and cycled 30 times at 94°C for 30 s, 55°C for 45 s, and 72°C for 1 min. Alleles were disrupted according to a published protocol (Baudin *et al.*, 1993). The ADH1/ADH2 double-deletion strain (YMT-1D) was made by mating the single-deletion strains, followed by tetrad dissection.

The double-deletion strain YMT-1D was transformed with the pRS411-ADH2, pRS415-ADH1, or pRS415-ADH_As vectors carrying wild-type ADH2, wild-type ADH1, and ancestral ADHs, respectively. Cells were grown in 2% glucose yeast minimal medium to mid-logarithmic phase and an OD_{600} value of 0.6. Extracts were applied to a Cibracon blue-coupled agarose column and protein was eluted with a gradient of NADH (0–200 μM). Fractions were pooled and applied to a Superdex 200 gel-filtration column to remove excess NAD^+(H) and to collect tetrameric ADH protein (Figure 14.3).

14.3.3 Ancestral vertebrate gene expressed in mammalian cells

Synthetic G-protein-coupled receptor genes are generally expressed in mammalian cells in tissue culture where pharmacological and cellular physiological effects can be correlated with structural changes introduced by mutation. In the case of ancestral rhodopsins, large quantities of the opsin apoprotein are produced in monkey kidney cells by transfection, where transcription is under the control of the human adenovirus major-late promoter, or in stable cell lines (Chang *et al.*, 2002). The apoprotein in the plasma membrane is regenerated with the chromophore 11-*cis*-retinal to

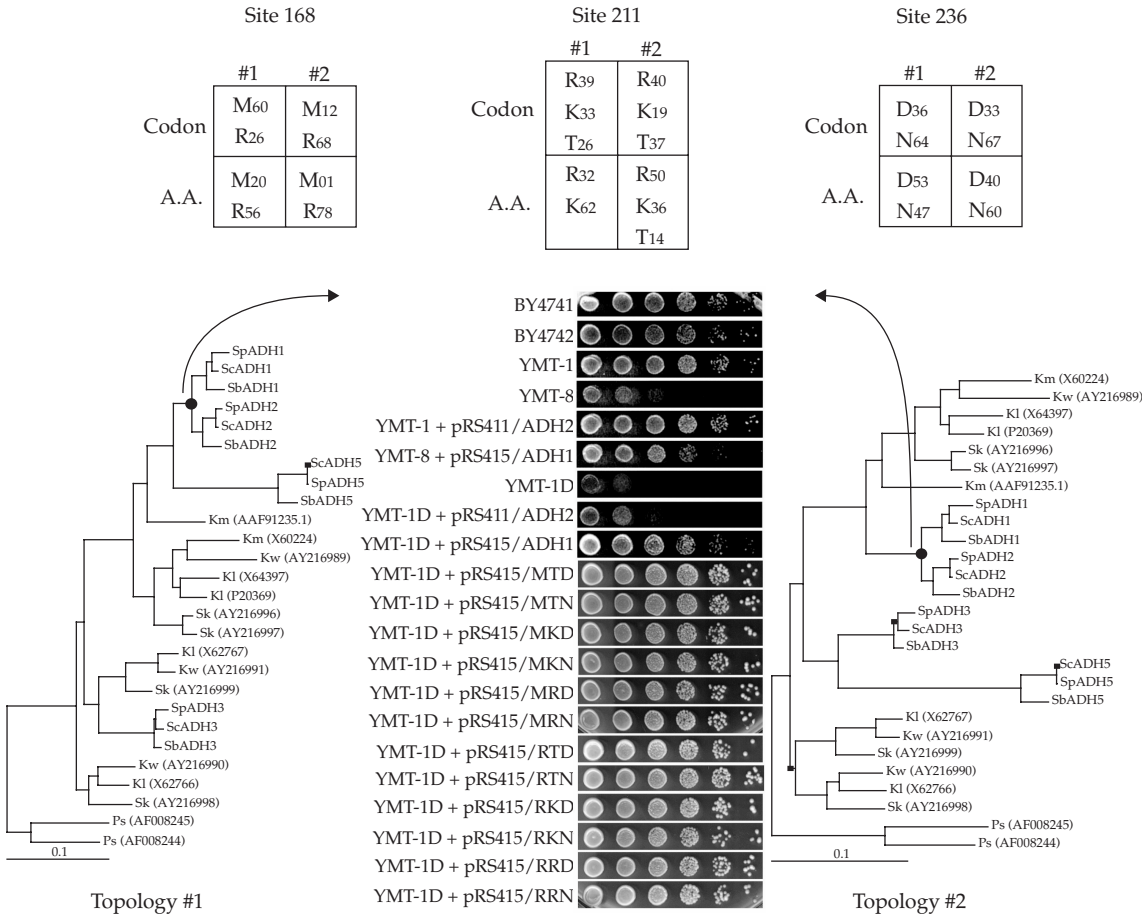

Figure 14.3 Accounting for ambiguity in reconstructed alcohol dehydrogenases (ADHs). The top panel presents differences in inferred residues for ancient ADH when considering analogous nodes from two competing phylogenies (topologies 1 and 2) and different evolutionary analyses (codon- and amino acid- (A.A.) based analyses and their posterior probabilities for individual residues). Three sites were considered ambiguous and 12 variants were synthesized to account for this ambiguity. The ability of ancient alcohol dehydrogenase variants to rescue knockouts of endogenous ADH1 and/or ADH2 in yeast S. cerevisiae as measured by a pin-stamp assay is presented in the middle of the bottom panel. BY4741 and BY4742 are isogenic strains discussed in the text. YMT-1 and YMT-8 are knockouts of ADH2 and ADH1 respectively. YMT-1D is a double knockout of ADH1 and ADH2. Vector pRS411 carries a wild-type ADH2 gene, while vector pRS415 carries either a wild-type ADH1 gene or the ancestral variants of the ADH1/ADH2 ancestral node highlighted on the phylogenies. As seen, the double knockout does not grow well and wild-type ADH2 does not rescue the wild-type phenotype. The double knockout is, however, rescued by wild-type ADH1 and all ancestral ADHs. Reprinted from Gaucher *et al.* (2003) Inferring the paleoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* **425**: 285–288.

form rhodopsin. The recombinant rhodopsin is then solubilized with detergent treatment and purified using an affinity adsorption method (see Chapter 15).

14.3.4 Ancestral monkey gene expressed in bacterial cells

EDN (or RNase 2) and eosinophil cationic protein (ECP; or RNase 3) are ribonucleases belonging to the RNase A family and are hypothesized to have arisen from a duplication event about 31 million years ago in the evolutionary lineage of hominoids and Old World monkeys (Zhang and Rosenberg, 2002). Since the inferred ancestral protein representing the common ancestor of EDNs and ECPs can be assayed *in vitro*, and ribonucleases can be heterologously translated in bacteria, the ancient primate genes can be expressed in *E. coli*.

Ancestral EDN/ECP sequences were constructed by site-directed mutagenesis using the owl monkey EDN sequence as the backbone (Zhang and Rosenberg, 2002). The ancient sequences were cloned into the pFCTS bacterial expression vector (IBI). The pFCTS vector adds the octapeptide DYKDDDDK (the FLAG epitope) to the recombinant protein, which permits its isolation and detection using the M2 anti-FLAG monoclonal antibody. These authors have demonstrated that the FLAG octapeptide does not interfere with the folding or catalytic activity of recombinant ribonucleases. Recombinant proteins were isolated from bacterial cultures after induction with isopropyl β-D-thiogalactoside (IPTG; 1 mM for EDNs, and 1 μM for human ECP). After harvest and sucrose lysis (EDNs) or harvest and cell lysis by freeze–thaw and sonication (human ECP), recombinant proteins were concentrated and isolated by M2 monoclonal antibody–agarose affinity chromatography (IBI).

14.4 Functional assays

Biological diversity is generally achieved through modifications to chromosomal architecture, gene expression, and protein structure. Transposition of mobile genetic elements is one mechanism that leads to chromosomal rearrangements which can subsequently generate novel proteins via domain-swapping or novel expression patterns of endogenous genes by modifying regulatory elements. Gene-expression patterns and protein structures can also be modified by point mutations to genomic DNA. For protein structure, mutations in the primary structure (amino acid sequence) are propagated through the secondary structure (helices, strands, etc.) and up to the tertiary structure (protein folds) where biological function is carried out. Here, function is based on the biomolecule's ability to contribute to the organismal success of its host through physical properties such as thermodynamics and protein stability, substrate recognition and catalysis, cofactor binding, and protein–protein interactions, among others. Ancestral sequence reconstructions and resurrections provide an opportunity to determine the specific physical behaviors of biomolecules that are responsible for functional divergence and thus biological diversity (Table 14.1).

The activation of mobile genetic elements such as transposons and retroposons throughout a genome can have a major impact on chromosomal architecture. Although mostly inactive during a species' lifetime, these elements can be activated during prolonged cellular stress and lead to mutant phenotypes that contribute to biological innovations such as speciation events. To date, three resurrection studies have verified that inactive, or dead, transposons and retroposons found in extant organisms (fish, frog, and mouse) were indeed active in their ancestral forms (Adey *et al*., 1994; Ivics *et al*., 1997; Miskey *et al*., 2003). These studies monitored transcription/expression and transposition of the ancient elements using a chloramphenicol acetyltransferase assay and ability of a neomycin-resistance gene to integrate into a host genome.

The contributions of protein structure and behavior to organismal fitness are more discernable than the contributions of mobile genetic elements, and this is reflected by ancestral resurrection studies. These studies often attempt to determine the behavior of an ancestral biomolecule whose descendents exhibit diverse biological behaviors, whereby diversity is generated from the divergence of paralogous and/or orthologous sequences.

Functional divergence of substrate-binding and catalysis of *paralogous* proteins has stimulated many resurrection studies. For instance, the kinetic interconversion of ethanol and acetaldehyde by alcohol dehydrogenase paralogs in yeast has been the focus of one study (Thomson *et al*., 2005), whereas the binding of different substrates to an ancient steroid receptor in invertebrates (as measured by the ability of steroids such as estradiol, progesterone, testosterone, and others to active an ancient receptor) and mammalian proteases (as measured by the ability of ancient chymases to degrade various angiotensins) have been the foci of others (Chandrasekharan *et al*., 1996; Thornton *et al*., 2003). Further, DNA-binding properties of ancient *Pax* paralogs in metazoans (as measured against various DNA substrates), optimal activation of ancient zebrafish visual pigment paralogs

Table 14.1 Examples of heterologous expression and functional assays of resurrected genes.

Gene family	Expression system	Reference	Functional assay
Digestive ribonucleases	Bacterial expression, assay *in vitro*	Stackhouse *et al.* (1990)	Single-stranded RNA hydrolysis, thermostability
Digestive ribonucleases	Bacterial expression, assay *in vitro*	Jermann *et al.* (1995)	Single- and double-stranded RNA hydrolysis, thermostability
Lysozyme	Bacterial expression, assay *in vitro*	Malcolm *et al.* (1990)	Bacterial cell-wall degradation, thermostability
L1 retroposons in mouse	Mouse cells	Adey *et al.* (1994)	Chloramphenicol acetyltransferase assay
Chymase proteases	Insect cell expression, assay *in vitro*	Chandrasekaran *et al.* (1996)	Angiotensin degradation
Tc1/mariner transposons	Human HeLa cells	Ivics *et al.* (1997)	Transposition of neomycin-resistance gene
Immune RNases	Bacterial expression, assay *in vitro*	Zhang and Rosenberg (2002)	RNA hydrolysis and ability to reduce viral infectivity
Pax transcription factors	Drosophila, in vivo	Sun *et al.* (2002)	DNA-binding assay and in vivo influence on development
SWS1 visual pigment	Monkey cell expression, assay *in vitro*	Shi and Yokoyama (2003)	Absorption spectra
Vertebrate rhodopsins	Monkey cell expression and assay	Chang *et al.* (2002)	Absorption spectra
Fish opsins (blue, green)	Monkey cell expression, assay *in vitro*	Chinen *et al.* (2005)	Absorption spectra
Steroid hormone receptors	Hamster cells	Thornton *et al.* (2003)	Transcriptional activation of estrogen-response element
Yeast alcohol dehydrogenase	Yeast	Thomson *et al.* (2005)	Binding kinetics to ethanol, acetaldehyde and cofactors
Green fluorescent proteins	Bacterial expression, assay *in vitro* and *in vivo*	Ugalde *et al.* (2004)	Color emission spectra
Isocitrate dehydrogenase	Bacteria	Zhu *et al.* (2005)	Cofactor-binding kinetics and chemostat competition
Isopropylmalate dehydrogenase	Bacterial expression, assay *in vitro*	Miyazaki *et al.* (2001)	Thermostability
Isocitrate dehydrogenase	Bacterial expression, assay *in vitro*	Iwabata *et al.* (2005)	Thermostability
Elongation factors	Bacterial expression, assay *in vitro*	Gaucher *et al.* (2003)	Thermostability

(as measured by spectral absorption), and ribonuclease paralogs in monkeys (as measured by RNA hydrolysis and ability to reduce viral infectivity) have been studied (Sun *et al.*, 2002; Chinen *et al.*, 2005; Zhang, 2006). In each case, resurrection of the last common ancestor preceding the duplication event produced an active ancestral biomolecule whose behavior was consistent within the context of the descendent sequences.

Functional divergence of substrate binding and catalysis among *orthologs* has also inspired ancestral resurrection studies. For instance, the ability of ancient artiodactyl ribonucleases to act on different substrates has been analyzed (measured by the ability to hydrolyze single- and double-stranded RNA), as has the behavior of ancient fluorescent proteins (measured by color emission spectra; Jermann *et al.*, 1995; Ugalde *et al.*, 2004). In addition to the study of paralogous visual pigments, ancestral forms of the orthologous vertebrate visual pigments rhodopsin and SWS1 have been separately resurrected and their

absorption spectra determined (Chang *et al*., 2002; Shi and Yokoyama, 2003).

Several other resurrection studies of ancient sequences inferred from orthologous genes have exploited the notion that protein function is dictated by protein folding and that tertiary-structure-based assays provide an opportunity to monitor changes in protein function. For instance, thermostability profiles of ancient isopropylmalate dehydrogenase, isocitrate dehydrogenase, and elongation factor proteins have been determined to infer the environmental temperature of ancient bacteria (Miyazaki *et al*., 2001; Gaucher *et al*., 2003; Iwabata *et al*., 2005). Further, the thermodynamic properties of ancient ribonucleases and lysozymes were correlated to the enzymatic properties of these ancient proteins (Malcolm *et al*., 1990; Jermann *et al*., 1995).

The above examples highlight the broad biomolecular properties that can be measured for resurrected DNA and protein. Substrate recognition, enzymatic catalysis, protein–protein/protein–DNA interactions, stabilization of transition states through cofactor binding, thermodynamics and protein stability, gene expression, genetic transposition, fluorescence via intramolecular interactions, and protein activation in response to visible-light stimulation are some of the properties *in vitro* and *in vivo* that can be exploited to determine ancient biomolecular behaviors.

14.5 Discussion

The utility of ancestral resurrection studies relies on an ability to connect ancient sequences to ancient molecular behaviors. This connection often consists of determining which of two extant molecular properties was present in an ancestral biomolecule. For instance, one can ask whether an ancient alcohol dehydrogenase converted ethanol to acetaldehyde (similar to the modern ADH1 enzyme) or whether it converted acetaldehyde to ethanol (similar to the modern ADH2 enzyme). Under this condition, an ancient duplication event would have resulted in neofunctionalization whereby one paralog retained the general molecular behavior of the ancestral form, while the other paralog acquired a novel biochemical behavior. Such a scenario is also consistent with other resurrection studies including, but not limited to, isocitrate/isopropylmalate dehydrogenase, steroid hormone receptors, and chymases.

Alternatively, a condition that involves subfunctionalization may be required to explain the evolution of ribonuclease paralogs. Here, modern paralogs display lineage-specific properties against a variety of RNA substrates. Conversely, the ancestral forms of ribonucleases display broad specificities in terms of hydrolysis of single- and double-stranded RNA, as well as immunosuppression. It appears that ancient ribonucleases were generalists in regards to RNA hydrolysis and that formation of paralogs though ancient duplication events allowed the generalists to become more specialized over time (honing some molecular behaviors, while shedding others). Since the precise selective pressures that shaped ancient ribonucleases are unknown, we cannot definitively know whether the broad molecular behaviors of ancient ribonucleases were simply promiscuous behaviors or were rather guided by natural selection. The inferences drawn from these ancient behaviors can, however, be supported by other lines of evidence such as those drawn from chemical theory, geology, animal physiology, and paleontology.

A third condition consists of resurrecting so-called dead genes that were once active but whose transcription/translation was subsequently abolished, through either mutations in regulatory elements and/or nonsense mutations in the coding regions of the gene. Examples of this condition include studies on ancient promoters and inactivated transposons. For instance, the genomes of various vertebrate species contain inactivated transposons with stop codons in unique positions but whose sequences are highly similar—suggesting that these elements were active in recent history. Reconstruction and resurrection of these elements demonstrated that the ancestral forms of the biomolecules were indeed active, as measured by their ability to transpose throughout a host genome. In a similar study, the inactivated promoter of an extant murine

retroposon was determined to be active in its ancestral form. In all, these studies distinctively fulfill the resurrection moniker.

A fourth condition consists of inferring an ancient sequence whose orthologous descendents encode a large range of molecular phenotypes. For instance, visual pigments such a rhodopsins exhibit absorption maxima between 480 and 510 nm in modern vertebrates. Although this quantitative difference may be negligible for an enzymatic reaction, the difference in absorption maxima for rhodopsins can determine whether a visually acute terrestrial animal is active during the day, dusk/dawn, or night. For aquatic animals, this difference can determine whether an organism lives in shallow-, mid-, or deep-range waters. As such, the absorption maximum of an ancestral rhodopsin can suggest a specific environmental niche for the host of the ancestral protein. Fluorescent proteins provide another example of this fourth condition since modern orthologs display a broad range of emission wavelengths (from red to green to non-fluorescent blue). Similarly, the thermostabilities of modern proteins exhibit a large range of temperature optima (5°C for proteins from psychrophiles to 100°C for proteins from hyperthermophiles) and these are linearly correlated to the environmental temperature of their respective hosts. As such, thermostability of ancient proteins is a good proxy to infer niche-specific temperatures of ancient organisms.

Functional divergence among homologous sequences is the *raison d'être* that supports ancestral reconstruction and resurrection studies. There is clearly little utility in resurrecting ancient biomolecules in the absence of functional divergence and biological diversity at the molecule level. The discussion above presents just a few of evolutionary scenarios that generate this functional divergence. These include neo- and subfunctionalization after gene duplication, gene inactivation and pseudogenization, and species-specific behaviors among orthologous sequences. For each scenario, an experiment in paleogenetics has provided insight into the evolutionary path that produced functional divergence.

14.6 Conclusions

The overview presented in this chapter reflects the diversity of approaches used for gene synthesis, heterologous expression, and functional assays of ancestral genes. We anticipate that these approaches will be further developed as the field itself progresses. For instance, a greater understanding of the statistical methods (and their associated biases) employed to infer ancestral sequences will require more sophisticated approaches to account for site-specific ambiguity during synthesis of the ancestral genes themselves (see Chapter 15). We also anticipate that functional assays will become more sophisticated as molecular reconstructionists try to answer higher-order questions regarding evolutionary paths and trajectories (Lunzer *et al.*, 2005; Weinreich *et al.*, 2006). These include experiments attempting to replay the molecular tape of life by integrating ancestral genes in modern organisms and allowing them to evolve in the laboratory. Further, this sophistication will be achieved as the field enters into an evolutionary synthetic biology whereby the evolutionary models and approaches of ancestral sequence reconstruction guide the engineering requirements of synthetic biology (see Chapter 2).

References

Adey, N.B., Tollefsbol, T.O., Sparks, A.B., Edgell, M.H., and Hutchison, C.A. (1994) Molecular resurrection of an extinct ancestral promoter for mouse L1. *Proc. Natl. Acad. Sci. USA* **91**: 1569–1573.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993) A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **21**: 3329–3330.

Chandrasekharan, U.M., Sanker, S., Glynias, M.J., Karnik, S.S., and Husain, A. (1996) Angiotensin II-forming activity in a reconstructed ancestral chymase. *Science* **271**: 502–505.

Chang, B.S., Jonsson, K., Kazmi, M.A., Donoghue, M.J., and Sakmar, T.P. (2002) Recreating a functional ancestral archosaur visual pigment. *Mol. Biol. Evol.* **19**: 1483–1489.

Chinen, A., Matsumoto, Y., and Kawamura, S. (2005) Reconstitution of ancestral green visual pigments of zebrafish and molecular mechanism of their spectral differentiation. *Mol. Biol. Evol.* **22**: 1001–1010.

Gaucher, E.A., Thomson, J.M., Burgan, M.F., and Benner, S.A. (2003) Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* **425**: 285–288.

Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. (1997) Molecular reconstruction of Sleeping beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**: 501–510.

Iwabata, H., Watanabe, K., Ohkuri, T., Yokobori, S., and Yamagishi, A. (2005) Thermostability of ancestral mutants of *Caldococcus noboribetus* isocitrate dehydrogenase. *FEMS Microbiol. Lett.* **243**: 393–398.

Jermann, T.M., Opitz, J.G., Stackhouse, J., and Benner, S.A. (1995) Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374**: 57–59.

Lunzer, M., Milter, S.P., Felsheim, R., and Dean, A.M. (2005) The biochemical architecture of an ancient adaptive landscape. *Science* **310**: 499–501.

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

Miskey, C., Izsvak, Z., Plasterk, R.H., and Ivics, Z. (2003) The Frog Prince: a reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells. *Nucleic Acids Res.* **31**: 6873–6881.

Miyazaki, J., Nakaya, S., Suzuki, T., Tamakoshi, M., Oshima, T., and Yamagishi, A. (2001) Ancestral residues stabilizing 3-isopropylmalate dehydrogenase of an extreme thermophile: experimental evidence supporting the thermophilic common ancestor hypothesis. *J. Biochem. (Tokyo)* **129**: 777–782.

Nambiar, K.P., Stackhouse, J., Stauffer, D.M., Kennedy, W.P., Eldredge, J.K., and Benner, S.A. (1984) Total synthesis and cloning of a gene coding for the ribonuclease S protein. *Science* **223**: 1299–1301.

Pauling, L. and Zuckerkandl, E. (1963) Chemical paleogenetics molecular restoration studies of extinct forms of life. *Acta Chem. Scand.* **17**: S9–S16.

Sassi, S.O., Benner, S.A., and Gaucher, E.A. (2007) Molecular paleosciences. Systems biology from the past. In *Advances in Enzymology and Related Areas of Molecular Biology: Protein Evolution*, vol. 75, pp. 1–132 (Toone, E., ed.). Wiley, Chichester.

Shi, Y. and Yokoyama, S. (2003) Molecular analysis of the evolutionary significance of ultraviolet vision in vertebrates. *Proc. Natl. Acad. Sci. USA* **100**: 8308–8313.

Stackhouse, J., Presnell, S.R., Mcgeehan, G.M., Nambiar, K.P., and Benner, S.A. (1990) The ribonuclease from an extinct bovid ruminant. *FEBS Lett.* **262**: 104–106.

Sun, H., Merugu, S., Gu, X., Kang, Y.Y., Dickinson, D.P., Callaerts, P., and Li, W.H. (2002) Identification of essential amino acid changes in paired domain evolution using a novel combination of evolutionary analysis and *in vitro* and in vivo studies. *Mol. Biol. Evol.* **19**: 1490–1500.

Thomson, J.M., Gaucher, E.A., Burgan, M.F., De Kee, D.W., Li, T., Aris, J.P., and Benner, S.A. (2005) Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat. Genet.* **37**: 630–635.

Thornton, J.W. (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**: 366–375.

Thornton, J.W., Need, E., and Crews, D. (2003) Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**: 1714–1717.

Ugalde, J.A., Chang, B.S., and Matz, M.V. (2004) Evolution of coral pigments recreated. *Science* **305**: 1433.

Weinreich, D.M., Delaney, N.F., Depristo, M.A., and Hartl, D.L. (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**: 111–114.

Zhang, J. (2006) Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat. Genet.* **38**: 819–823.

Zhang, J.Z. and Rosenberg, H.F. (2002) Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. *Proc. Natl. Acad. Sci. USA* **99**: 5486–5491.

Zhu, G.P., Golding, G.B., and Dean, A.M. (2005) The selective cause of an ancient adaptation. *Science* **307**: 1279–1282.