Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Impact of ancestral sequence reconstruction on mechanistic and structural enzymology

Callum R. Nicoll[1], Marta Massari[1], Marco W. Fraaije[2], Maria Laura Mascotti[2,3] and Andrea Mattevi[1]

Abstract

Ancestral sequence reconstruction (ASR) provides insight into the changes within a protein sequence across evolution. More specifically, it can illustrate how specific amino acid changes give rise to different phenotypes within a protein family. Over the last few decades it has established itself as a powerful technique for revealing molecular common denominators that govern enzyme function. Here, we describe the strength of ASR in unveiling catalytic mechanisms and emerging phenotypes for a range of different proteins, also highlighting biotechnological applications the methodology can provide.

Addresses

[1] Department of Biology and Biotechnology Lazzaro Spallanzani, University of Pavia, Via Ferrata 9, 27100, Pavia, Italy

[2] Molecular Enzymology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747, AG Groningen, the Netherlands

[3] IMIBIO-SL CONICET, Facultad de Química Bioquímica y Farmacia, Universidad Nacional de San Luis, Ejército de los Andes 950, D5700HHW, San Luis, Argentina

Corresponding author: Mattevi, Andrea (andrea.mattevi@unipv.it) (Fraaije M.W.) (Mattevi A.)

Current Opinion in Structural Biology 2023, **82**:102669

This review comes from a themed issue on **Catalysis and Regulation (2023)**

Edited by **Emily Parker** and **Christopher Schofield**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.sbi.2023.102669

Keywords

Ancestral sequence reconstruction, Enzyme evolution, Enzyme biotechnology, Mechanistic enzymology, Protein stability.

Introduction

Ancestral sequence reconstruction, ASR, also known as ancestral protein reconstruction, is a molecular evolution technique that infers the sequences of extinct proteins using probabilistic methods including maximum likelihood and Bayesian inference. Whilst not being able to determine the true sequence of an ancestral protein, ASR captures the changes underlying a protein phenotype occurring across sequence-space as a function of time. Experimental characterisation of various ancestors enables one to understand the determinants permitting a given biochemical function [1–4]. Pauling and Zuckerkandl envisioned that one day it would be possible to routinely synthesise ancestral proteins that existed in extinct organisms [5]. At the time, bioinformatic technologies and sequence databases were limited restricting the phylogenetic histories that could be mapped. Since the late 20th century there has been a massive uplift in computing power and the amalgamation of sequence databases has institutionalised a new setting for ASR. Indeed, to date a plethora of ancestral proteins have been resurrected and characterised [6–12].

Reliable reconstruction of an ancestral sequence hinges on a robust phylogenetic tree that frames the biochemical setting for the to-be undertaken research and the phenotype under scrutiny [13-18*]. The resurrection component concerns the wet lab with the sequences synthesised, expressed, purified, and characterised. Here, several ancestral proteins, reflecting the evolution of a given phenotype, are evaluated and their sequence determinants are described.

ASR research over the last decade has seen a revolution in terms of its use in biochemical investigations and its potential application in industry. The development and consolidation of cheap DNA synthesis, powerful new experimental techniques (*i.e.*: mass photometry, expression systems, kinetics) and *in silico* advances (simulations, Alphafold) have strongly contributed to the flourishing of ASR. In this review, we discuss selected case-studies highlighting the prowess that ASR has recently brought in modern biochemical investigations and biotechnological applications for structural and enzymological studies (Table 1).

Fine tuning of an oxygenating flavin intermediate for xenobiotic metabolism

Detoxification is an essential component of secondary metabolism. A wealth of detoxifying enzymes has

Table 1

ASR mechanistic and structural enzymology studies. The most recent investigations (2019-to date) provide ground-breaking insight into mechanistic or structural enzymology and are summarised here.

Enzyme/protein	Mechanistic/structural enzymology concept
Vertebrates chemokines [9*]	• Fold switching
Bacterial & eukaryotic family 1-glycosidases [26*]	• Vestigial heme-mediated catalysis • Flexible structure with unexpected cofactor binding • Allosteric regulation
Opisthokont Ser−Thr protein kinase [7]	• Emergence of allosteric regulation
Bacterial solute binding proteins [8]	• Origins of catalysis
Bacterial PETases [37*]	• Emergence from cutinases • Catalytic features • Substrate (PET polymer) specificity
Prokaryotic haloalkane dehydrogenases [32]	• Catalytic promiscuity
Vertebrate haemoglobin [46]	• Origin of cooperativity • Oxygen binding • Determinants of multimerisation
Mammalian flavin-containing monooxygenases (FMOs) [22*]	• Substrate/product entrance/exit • Catalytic diversity
Rubiscos [30*-31**]	• Catalytic traits in C3 plants • Dependency on an accessory subunit
Fe/Mn superoxide dismutases (SODs) [27*]	• Metal dependency for activity
Vertebrate cytochrome P450 family 1 [18*]	• Substrate specialisation
tRNA nucleotidyltransferase [47*]	• K_M versus k_{cat} compensation
Bacterial Cas9 [12]	• Transition nickase to double-strand breaking activity

evolved to breakdown and remove a range of xenobiotics [19*], including the flavin-containing monooxygenases, FMOs [20,21]. Humans possess five different FMO members with four performing heteroatom oxygenations and one performing Baeyer−Villiger oxidations. Both reactions require the production of a reactive C4a-(hydro)peroxyflavin intermediate for catalysis. The protonation state of the intermediate has been postulated to dictate the resulting catalytic trajectory; the protonated version enables heteroatom oxygenation and the deprotonated version is responsible for Baeyer−Villiger activity [22]. Phylogenetic analysis reveals that the 'explosion' of the FMO genes took place at the dawn of the tetrapods (four-limbed creatures) approximately 300 million years ago (Mya). The first FMO to emerge was FMO5, which is responsible for the Baeyer−Villiger monooxygenase activity. The second branch corresponded to the ancestor of all other FMOs, 1−4, that are expected to pursue heteroatom oxygenation, with FMO4 emerging first, followed by FMO2, and then FMO1 and FMO3 (Figure 1a). Mascotti and colleagues reconstructed four different ancestral intermediates to inform on the changing phenotypes: tAncFMO1-5 (t refers to tetrapod ancestry and 1−5 mean ancestor to FMOs 1 through to 5), tAncFMO5, tAncFMO1-4 and tAncFMO1-3 (see Figure 1a) [23*]. Using a set of canonical substrates consisting of ketones (Baeyer−Villiger substrates) and molecules possessing soft sulphur- and nitrogen-nucleophilic centres (heteroatom-containing substrates) the different ancestors were screened for activity. The first ancestor, tAncFMO1-5, exhibited mixed functionality and could turnover, albeit in low yields, most substrates. This trend was also observed for tAncFMO5, but with slightly diminished heteroatom oxygenations. However, ancestors tAncFMO1-4 and tAncFMO1-3 showed significantly higher heteroatom oxygenation capabilities with ablation of Baeyer−Villiger oxidation activity (Figure 1b). This finding suggests that the gene duplication event produced a highly specialised heteroatom oxygenase that underwent further gene duplications to optimise oxygenation events.

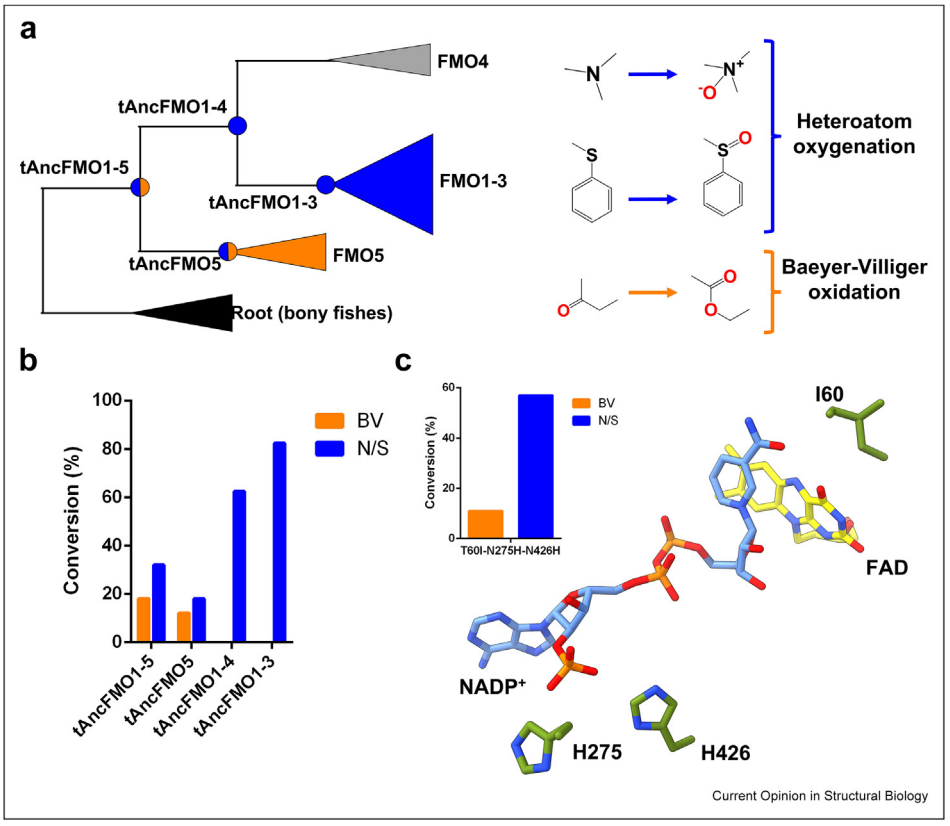
With the switch in phenotype being pinpointed to the tAncFMO1-5 and tAncFMO1-4 ancestral states, the authors evaluated the differences between the two proteins (45 residues in total) [23*]. Using a rational approach considering sequence and structure elements, they reduced the number of relevant residues down to 16 and generated several tAncFMO1-4 mutants encompassing increasing numbers of substitutions aiming to recapitulate the Baeyer−Villiger monooxygenase activity. After several mutational rounds, they observed that just four residues were enough to confer Baeyer−Villiger activity. These were further dissected into single, double and triple combinations. Neither of the single point mutants introduced the switch whereas only the combination of three of them restored the Baeyer−Villiger oxidation (Figure 1c). Inspecting the structural positions of these residues shows two distal

regions. The first is nestled close to the isoalloxazine ring of the FAD. The second, on the periphery of the protein core, is positioned in the vicinity of the adenosine motif of NADPH (Figure 1c). Mutations in both regions introduce subtle, yet relevant, differences among the vast interactions that surround the ligands engaging with the oxygenating C4a-(hydro)peroxyflavin intermediate — $NADP^+$ and FAD. In this scenario, likely, both the FAD and $NADP^+$ cofactors relay the altered electronic properties of the newly introduced residues towards the oxygenating intermediate. This consequently changes its available protonation states and thereby, its mode of function. The evolutionary mechanism behind this process has been described as epitasis relayed by the ligand; it has been recognised as one of the most complex catalytic networks to be established [24–26].

Cofactors on the rise

Most enzymes employ organic and metal cofactors as essential elements of their catalytic apparatus. ASR is becoming a potent tool to trace the emergence and evolution of the cofactor-binding sites, often providing mechanistic insight about the modulation of the cofactor reactivity by the protein environment. Sanchez-Ruiz et al. [27*] serendipitously found that an old ancestor of family-1 glycosidases was endowed with heme binding. This is surprising because none of the modern enzymes were previously reported to bind this cofactor. Moreover, family-1 glycosidases feature a $(\beta/\alpha)_8$-barrel and heme binding is rare among the proteins displaying this folding topology. Nonetheless, heme incorporation is unlikely to be a spurious artefact of the resurrection protocol. First, the binding site is partly conserved among the extant family-1 glycosidases and

Figure 1



Evolution of the flavin-containing monooxygenases in tetrapods. (**a**) Scheme for the phylogenetic tree of the FMO family in tetrapods. The tetrapod ancestral states investigated are shown with circles. The clade possessing Baeyer–Villiger monooxygenase activity is shown in orange (FMO5) whilst the heteroatom oxygenating one is shown in blue (FMO1-3). The uncharacterised FMO4 clade is shown in grey with the root of the tree, composed of fish FMOs, shown in black. Examples of both heteroatom and Baeyer–Villiger oxidations are shown. (**b**) Substrate conversions across ancestral FMOs. The bar chart portrays an average percentage conversion for Baeyer–Villiger substrates versus heteroatom-containing substrates. (**c**) Underlying residues responsible for introducing Baeyer–Villiger activity. Residues H275, H246 and I60, shown in green, are those found in the tAncFMO1-5 backbone that were introduced into tAncFMO1-4. Cofactor FAD and coenzyme $NADP^+$ are shown in yellow and cornflower blue, respectively. The inset bar chart portrays an average percentage conversion for Baeyer–Villiger substrates versus heteroatom-containing substrates.

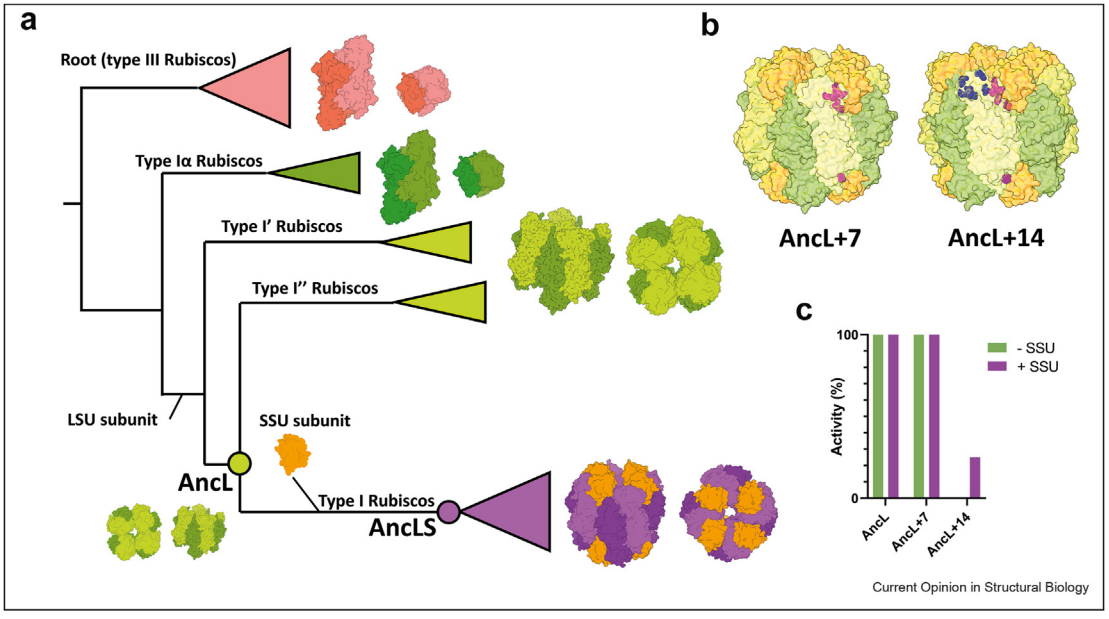
the authors retrospectively found that weak heme binding is present in several extant family members. Moreover, removal of the heme leads to drastic reduction of the enzymatic activity of the heme-binding ancestor and is associated with reduced protein flexibility. In line with these findings, recently, Sanchez-Ruiz and colleagues observed that an unrelated TIM-barrel protein was able to sequester free heme using a flexible region of the polypeptide to promote a peroxidase mode of function and reduce the half-life of free heme in solution [28]. Collectively, these findings portray that just a few mutations can craft a cofactor-binding site with a potential functional role.

Superoxidase dismutases (SODs) are widespread in nature. These metalloenzymes are classified based on their metal cofactor. The iron/manganese family members can be selective for iron or manganese or can be equally active with either metal (so-called cambialistic enzymes). Tawfik and co-workers [29*] reasoned that these SODs would represent an interesting system to study the evolution of metal dependency. They discovered that their oldest ancestors were likely cambialistic as the resurrected proteins display activity with both iron and manganese. Metal specificity has developed only later, depending on the evolutionary branches. Interestingly, the resurrected SOD belonging to the last universal oxygen ancestor, the first organism capable of metabolising oxygen, was strictly dependent on manganese and unable to use iron. From a geobiochemical standpoint, this would seem logical as iron bioavailability likely became limited in the early aerobic organisms that had to protect themselves from the potentially damaging Fenton reactions.

The evolution of specificity in form I Rubiscos

Rubiscos are integral to carbon assimilation in all aerobic phototrophs [30,31]. Rubiscos predate the emergence of oxygenic photosynthesis and evolved under anaerobic conditions. Molecular oxygen however drives a side reaction in Rubisco resulting in the formation of 2-phosphoglycolate that consequently interferes and reduces carbon-based metabolism. Among the Rubisco family, several forms have evolved including I, I', I'', Iα and III variants. Each possess unique molecular deterrents to dampen this side reaction [32*]. In type I Rubisco, the recruitment of a non-catalytic small subunit provides increased specificity for CO_2. To determine the roles of this additional protein chain, Hochberg and colleagues set out to uncover the molecular determinants giving rise to the interaction and heightened specificity using ASR, mass photometry, X-ray crystallography, Cryo-EM and mutagenesis studies [33**].

Figure 2

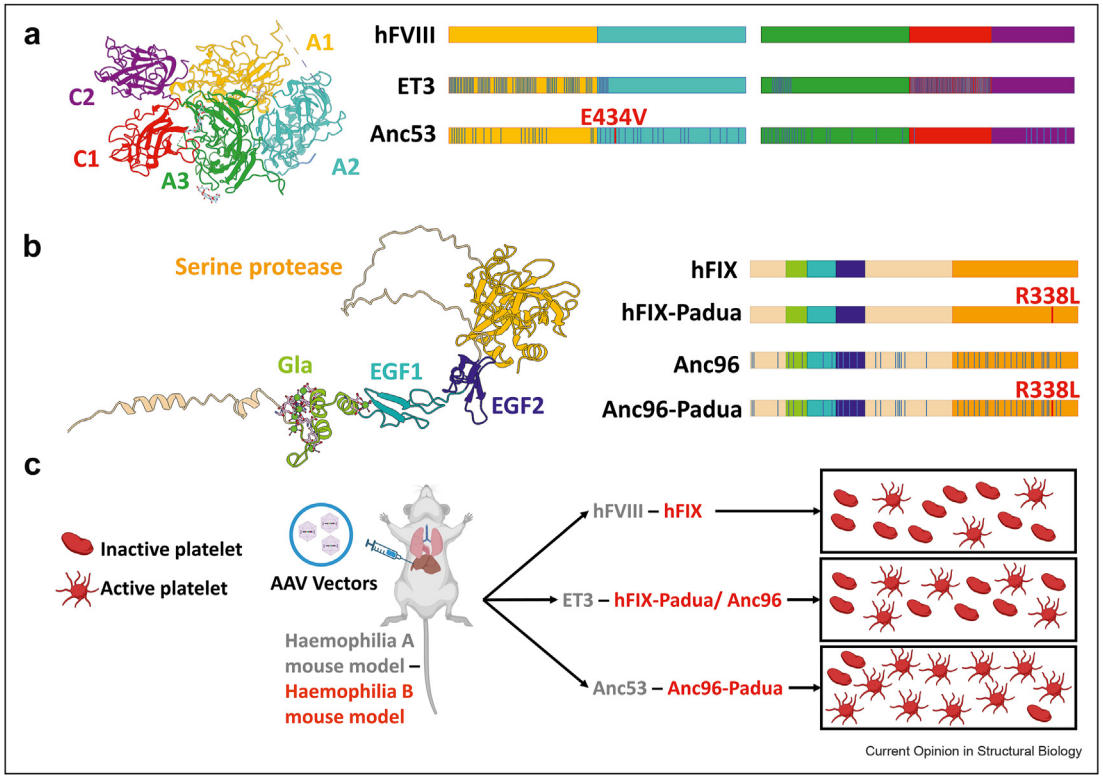


Evolution of form I Rubiscos. (**a**) A scheme of the phylogenetic tree of the Rubisco family. Each Rubisco family member is shown in its oligomeric forms perpendicular and parallel to the C4 rotation axis. Type III, Iα (PDB: 1YKW for structure representation), I′, I″ (PDB: 7QSY) and I Rubiscos (PDB: 7QVI) are shown in salmon, dark green, light green and purple, respectively. The SSU subunit is shown in orange with its introduction to function illustrated. The AncL and AncLS nodes are illustrated. (**b**) Location of changes found in AncL+7 and the additional changes added in AncL +14 are shown in magenta and blue, respectively. (**c**) Dependence and influence of activity of the SSU unit across the ancestral states, AncL, AncL + 7 and AncL + 14 are shown.

Form I Rubiscos are arranged into eight large subunits (LSU), meshed with eight non-catalytic small subunits (SSU), together making a 16-unit oligomer, L_8S_8. To understand how the components came together to create the large heterooligomeric structure, they reconstructed the ancestors appearing before and after the binding event, AncL and AncLS, respectively, as well as the ancestral small subunit, AncSSU (Figure 2a). They demonstrated that AncL and AncLS + AncSSU were catalytically competent systems, with the K_M of CO_2 decreasing for the latter and the catalytic efficiency increasing. AncSSU had no effect on the activity or yields of AncL and did not show any identification of binding when assessed through mass photometry or co-elution studies. Alternatively, in the absence of AncSSU, AncLS was enzymatically stunted. These findings illustrated that the SSU became rapidly essential towards function.

The authors documented the changes occurring between the ancestors, AncL and AncLS; after solving crystal structures of inhibitor bound AncL and co-crystallised AncLS + AncSSU, they could pinpoint the location of these changes with respect to the oligomerisation. Remarkably, only three residues were required to induce binding when introduced into AncL with a total of seven (hereafter denoted as AncL+7) giving rise to tight binding and the resulting L_8S_8 heterocomplex (Figure 2b). The authors further investigated the influence of AncSSU on the activity of AncL+7 (Figure 2c). They noted that the K_M of CO_2 decreased in the presence of AncSSU, and despite a slight reduction in catalytic rate, catalytic efficiency increased two-fold. Furthermore, AncL+7 exhibited specificity for CO_2, but significantly more so in the presence of AncSSU, thereby showing an increasing enhancement by both the residue changes and the recruitment of the

Figure 3



Ancestral sequence reconstruction for protein coagulation factors, VIII and IX. (**a**) Structural features and domain composition of FVIII. Left, 3D-structure of human FVIII (PDB: 3CDZ); right, schematic representation of the human FVIII, the human/porcine hybrid ET3 and the ancestral protein Anc53 sequences. Mutations compared to the human protein are represented as blue lines. The mutation E434V, shown in red, on Anc53 represents a single-point mutation performed on the ancestral sequence to avoid the binding of the hFVIII inhibitor MAb 4A4. (**b**) Structural features and domain composition of FIX. Left, 3D-view of human FIX (Alphafold model aligned with the crystal structure of the Gla domain (PDB: 1J35)); right, schematic representation of the wild-type FIX, the gain-of-function mutant (FIX-Padua [43]), the Anc96, and the Anc96-Padua sequences. Mutations compared to the human protein are represented as blue lines. The mutation R338L (in red) corresponds to the gain-of-function Padua mutation. (**c**) Illustration of liver-directed gene transfer through adeno-associated vector (AAV) on a murine model for haemophilia A (grey) or haemophilia B (red). Haemophilia A mice treated with Anc53 showed significantly higher coagulation activity than human FVIII and ET3. Haemophilia B mice treated with Anc96-Padua displayed higher activity than the human wild-type FIX, hFIX-Padua and Anc96, indicating a synergistic effect between ASR and the R338L mutation.

SSU. The addition of another residue, N170, present in AncLS, further increased specificity for CO_2 but only in the presence of AncSSU. Of note, none of these residues are within or proximal to the active site. After resolving crystal structures of AncL+7 and AncL+7+N170, in both L_8 and L_8S_8 forms, it became evident that SSU confers heightened catalysis through both allosteric and long-distance effects.

To reveal the residues that forfeit SSU dependency, the authors introduced seven residues into AncL+7 found to be neighbouring the L-S interface in the AncLS + AncSSU complex (AncL+14 in Figure 2b and c). They revealed that these residues were detrimental to AncL solubility and dependent on AncSSU co-expression. Further scrutiny of these individual residues revealed that one change, W437, present in AncLS, diminished protein solubility by approximately 75% in both AncL and AncL+7 ancestors, substantiating that this residue conferred SSU dependency.

ASR in biotechnology and biomedicine

The growing consensus of introduced thermostability in reconstructed proteins has promoted interest in its potential application in industry [34–39*] and biomedicine [40,41]. Harmer and colleagues [42**] recently proposed a 'survivor bias' model to explain this observation. Sequences that have evolved under selective pressure tend not to include destabilising residues. Thus, stabilising substitutions are overrepresented in reconstructed ancestors. These results build on the accumulating evidence collected by Goldstein and colleagues that maximum likelihood presents a bias towards more frequent amino acids, commonly referred to as the 'consensus effect' [43]. Their work suggests that Bayesian methods have a lower tendency to induce these thermostability-enhancing residues. Nevertheless, Arcus and colleagues reconstructed several ancestral 3-isopropylmalate dehydrogenases (LeuB), representing the last common ancestor of *Bacillus*, using both maximum likelihood and Bayesian inference methods and observed that both introduced heightened stability. Thus, both techniques can lead to thermostabilising artefacts [44,45]. Collectively, these findings showcase the underlining methodological attributes that frequently impart thermostability for ancestral proteins procured by ASR [46].

More valuable, however, concerns the retained functional characteristics as opposed to the improved physical stability. This is the case of the reconstruction and experimental validation of the coagulation factors FVIII and FIX from mammals with gene-therapy purposes (Figure 3). Recombinant human FVIII has been produced and used over the last thirty years to treat haemophilia A (Figure 3a) [47]. However, the human protein suffers from poor yields, stability and patients receiving FVIII infusions typically develop anti-FVIII antibodies against the recombinant protein, severely reducing the efficacy of the treatment. In 2017, Doering and collaborators explored a historical biochemistry approach to tackle this issue [41]. Interestingly, ancestral FVIII proteins showed significantly higher expression than the human FVIII and displayed a significant reduction in sensitivity to anti-human FVIII antibodies compared to the human protein (Figure 3a). Moreover, B-cell epitope prediction showed no differences between the human FVIII and ancestors, corroborating that ASR can be exploited for protein humanisation (Figure 3c).

As for FVIII deficiency, recombinant human FIX protein was produced for the treatment of haemophilia B patients and displayed similar issues as FVIII treatment (Figure 3c) [48]. Doering et al. [49*] generated AncFIX sequences, particularly focusing on two primate ancestors and, remarkably, exhibited around 10-fold higher activity than human FIX. Moreover, incorporation of a gain-of-function mutation [50] led to 5-fold higher production yields, 3-fold greater specific activity and lower 50% effective dose. Hence, ASR is proving to be a powerful and cost-effective method for drug improvement and, in combination with other enzyme engineering methods, is a useful tool in the biopharmaceutical field.

Fortune favours the few

In this opinion, we have exemplified how ASR can unravel enzymatic activities, protein–protein interactions, as well as showcasing novel biotechnological applications of ASR. Interestingly, these investigations reveal the significance of only a few residues can have on a phenotype when investigating complex systems. These *one-hit wonders* however seldom materialise entirely new phenotypes. Instead, they take the first step in evolutionary trajectories. For Rubisco, a single mutation, S147W, greatly reduced protein solubility that ushered in SSU, a partnership further enforced by allosteric and long-range interactions. Differently from these protein-protein interaction orientated studies, whereby one residue gave rise to a trajectory that was then reinforced with further changes, enzymatic activity in the FMOs has required fine-tuning. These minimum requirements were enabled by three residues; their distribution around the integral ligands invoked the partial electrostatic shift in the active site's microenvironment needed to incite Baeyer-Villiger oxidation. Again, in this example, long-range interactions play a fundamental role. Hence, we posit that, likely, the pillars for molecular evolution of multicomponent systems are grounded by a small number of residues that provide the hallmarks for new phenotypes (arguably preadaptation). These are then nourished through neighbouring and/or regional residues, with long distance substitutions playing an

instrumental part. An increasing number of examples are clearly highlighting this phenomenon [24–26,51–53*].

The future of ASR

A considerable hurdle that non-evolutionary biologists face when looking to using ASR is the methodology itself. The inference of a concrete phylogenetic tree including its various statistical approaches, appropriate rooting procedures and the introduction of alternative ancestors to assess the robustness of the analysis may seem daunting to a protein chemist. Nonetheless the increasing number of reviews and articles documenting how to carry out the technique are providing new opportunities to non-experts [1,13,14,18]. We envisage that soon, as phylogenetic analysis becomes more mainstream, researchers will include ASR as a complementary technique to substantiate their ongoing hypothesis towards a protein function. In this regard, biochemical research will routinely frame enzymatic function not just on one sequence but across a family as a function of time, rewording the statement from Theodosius Dobzhansky [54] to 'Nothing in Biochemistry Makes Sense Except in the Light of Evolution'.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The research on enzyme evolution and technology in the Authors' laboratories is supported by Fondazione Cariplo (grant 2020-0894), the European Union's Horizon 2020 research and innovation program under Grant Agreements 824017 (B-Ligzymes) and 847675 (COFUND project oLife). CN and MM are supported by the Associazione Italiana per la Ricerca sul Cancro (fellowships N° 28380 and 26648).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Merkl R, Sterner R: **Ancestral protein reconstruction: techniques and applications**. *Biol Chem* 2016, **397**:1–21.

2. Starr TN, Picton LK, Thornton JW: **Alternative evolutionary histories in the sequence space of an ancient protein**. *Nature* 2017, **549**:409–413.

3. Risso VA, Sanchez-Ruiz JM, Banu Ozkahn S: **Biotechnological and protein-engineering implications of ancestral protein resurrection**. *Curr Opin Struct Biol* 2018, **51**:106–115.

4. Schupfner M, Straub K, Busch F, Merkl R, Sterner R: **Analysis of allosteric communication in a multienzyme complex by ancestral sequence reconstruction**. *Proc Natl Acad Sci U S A* 2020, **117**:346–354.

5. Pauling L, Zuckerkandl E: **Chemical paleogenetics**. *Acta Chem Scand* 1963, **17**:S9–S16.

6. Voordeckers K, Brown CA, *et al.*: **Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication**. *PLoS Biol* 2012, **10**, e1001446.

7. Hadzipasic A, Wilson C, Nguyen V, *et al.*: **Ancient origins of allosteric activation in a Ser-Thr kinase**. *Science* 2020, **367**:912–917.

8. Kaczmarksi JA, Mahawaththa MC, *et al.*: **Altered conformation sampling along an evolutionary trajectory changes the catalytic activity of an enzyme**. *Nat Commun* 2020, **11**:5945.

9. * Dishman AF, Tyler RC, *et al.*: **Evolution of fold switching in a metamorphic protein**. *Science* 2021, **371**:86–90.
Metamorphic proteins are single polypeptide chains that undergo reversible and drastic folding changes to exert distinct functions. The authors studied the evolution of the chemokine XCL1, the only protein of its family that can switch to an alternate fold. By resurrecting several ancestors, they discovered that metamorphism evolved along XCL1 evolution through the disruption of a disulphide bridge and mutating residues involved in non-covalent interactions.

10. Yang G, Miton CM, Tokuriki N: **A mechanistic view of enzyme evolution**. *Protein Sci* 2020, **29**:1724–1747.

11. Thomson RES, Carrera-Pacheco SE, Gillam EMJ: **Engineering functional thermostable proteins using ancestral sequence reconstruction**. *J Biol Chem* 2022, **298**, 102435.

12. Alonso-Lerma B, Jabalera Y, Samperio S, *et al.*: **Evolution of CRISPR-associated endonucleases as inferred from resurrected proteins**. *Nat Microbiol* 2023, **8**:77–90.

13. Mascotti ML: **Resurrecting enzymes by ancestral sequence reconstruction**. *Enzyme Eng* 2022:2397.

14. Hochberg GKA, Thornton JW: **Reconstructing ancient proteins to understand the causes of structure and function**. *Annu Rev Biophys* 2017, **46**:247–269.

15. Harms MJ, Thornton JW: **Evolutionary biochemistry: revealing the historical and physical causes of protein properties**. *Nat Rev Genet* 2013, **14**:559–571.

16. Siddiq MA, Hochberg GKA, Thornton JW: **Evolution of protein specificity: insights from ancestral protein reconstruction**. *Curr Opin Struct Biol* 2017, **47**:113–122.

17. Gerlt JA, Babbitt PC: **Enzyme (re)design: lessons from natural evolution and computation**. *Curr Opin Chem Biol* 2009, **13**:10–18.

18. * Spence MA, Kaczmarski JA, Saunders JW, Jackson CJ: **Ancestral sequence reconstruction for protein engineers**. *Curr Opin Struct Biol* 2021, **69**:131–141.
An effective description of the different methods available to generate ancestral sequences and highlight elements of the approach that need attention when carrying out the analysis.

19. * Harris KL, Thomson RES, *et al.*: **Ancestral sequence reconstruction of a cytochrome P450 family involved in chemical defense reveals the functional evolution of a promiscuous, xenobiotic-metabolizing enzyme in vertebrates**. *Mol Biol Evol* 2022, **39**:msac116.
The resurrected vertebrate ancestors of a cytochrome P450 family are shown to be less active towards the xenobiotics and steroid substrates that are efficiently metabolized by the extant family members. For instance, only the younger ancestors are capable of efficient caffeine degradation mirroring the only recent emergence of plant caffein biosynthesis. These findings illustrate the constant-evolving activities of the xenobiotic-degrading enzymes.

20. Nicoll CR, Bailleul G, *et al.*: **Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs**. *Nat Struct Mol Biol* 2020, **27**:14–24.

21. Bailleul G, Nicoll CR, *et al.*: **Ancestral reconstruction of mammalian FMO1 enables structural determination, revealing unique features that explain its catalytic properties**. *J Biol Chem* 2021, **296**, 100221.

22. Sheng D, Ballou DP, Massey V: **Mechanistic studies of cyclohexanone monooxygenase: chemical properties of intermediates involved in catalysis**. *Biochemistry* 2001, **40**: 11156–11167.

23. * Bailleul G, Yang G, Nicoll CR, *et al.*: **Evolution of enzyme functionality in the flavin-containing monooxygenases**. *Nat Commun* 2023, **14**:1042.
Vertebrate FMOs are shown to derive from a common multi-functional ancestor that gave rise to more specialised extant enzymes through a series of gene duplications combined with mutations. The molecular mechanism behind functional divergence is unveiled.

24. Gong LI, Suchard MA, Bloom JD: **Stability-mediated epistasis constrains the evolution of an influenza protein**. *Elife* 2013, **2**, e00631.

25. Starr TN, Thornton JW: **Epistasis in protein evolution**. *Protein Sci* 2016, **25**:1204–1218.

26. Campitelli P, Swint-Kruse L, Banu Ozkan S: **Substitutions at non-conserved rheostat positions modulate function by rewiring long-range, dynamic interactions**. *Mol Biol Evol* 2021, **38**:201–214.

27. * Gamiz-Arco G, Gutierrez-Rus LI, *et al.*: **Heme-binding enables allosteric modulation in an ancient TIM-barrel glycosidase**. *Nat Commun* 2021, **12**:380.
Resurrection of an ancestor of a glycosidase family reveals an unexpected heme-binding site. The crystal structure shows that the heme is fully embedded in the protein, in the vicinity of the catalytic centre. Heme binding allosterically activates the enzyme by making the protein more rigid.

28. Gutierrez-Rus L, *et al.*: **Protection of catalytic cofactors by polypeptides as a driver for the emergence of primordial enzymes**. *Mol Biol Evol* 2023:msad126.

29. * Valenti R, Jabłonska J, *et al.*: **Characterization of ancestral Fe/Mn superoxide dismutases indicates their cambialistic origin**. *Protein Sci* 2022, **31**:e4423.
SODs of the iron/manganese family display variable metal selectivity. The authors trace the evolution of the cofactor dependency in relation with the onset of the oxygen-metabolizing organisms.

30. Flamholz AI, *et al.*: **Revisiting trade-offs between Rubisco kinetic parameters**. *Biochemistry* 2019, **58**:3365–3376.

31. Iñiguez C, *et al.*: **Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2 concentrating mechanisms**. *Plant J* 2020, **101**:897–918.

32. * Lin MT, Salihovic H, *et al.*: **Improving the efficiency of Rubisco by resurrecting its ancestors in the family Solanaceae**. *Sci Adv* 2022, **8**, eabm6871.
Ancestral Rubiscos from the *Solanaceae* family feature higher kinetic efficiency, demonstrating the power of ASR towards the generation of valuable enzyme variants for biotechnological purposes.

33. ** Schulz L, Guo Z, Zarzycki J, *et al.*: **Evolution of increased complexity and specificity at the dawn of form I Rubiscos**. *Science* 2022, **378**:155–160.
This paper reveals the evolutionary trajectory that rendered form I Rubiscos to be dependent on an ancillary subunit that drastically increased the efficiency of the carboxylation over the undesired oxygenation reaction.

34. Chalupkova R, Liskova V, *et al.*: **Light-emitting dehalogenases: reconstruction of multifunctional biocatalysts**. *ACS Catal* 2019, **9**:4810–4823.

35. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM: **Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases**. *J Am Chem Soc* 2013, **135**:2899–2902.

36. Devamani T, Rauwerdink AM, Lunzer M, Jones BJ, Mooney JL, Tan MAO, Zhang Z-J, Xu J-H, Dean AM, Kazlauskas RJ: **Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases**. *J Am Chem Soc* 2016, **138**:1046–1056.

37. Chen S, Xu Z, Ding B, *et al.*: **Big data mining, rational modification, and ancestral sequence reconstruction inferred multiple xylose isomerases for biorefinery**. *Sci Adv* 2023, **9**, eadd8835.

38. Livada J, Vargas AM, Martinez CA, Lewis RD: **Ancestral sequence reconstruction enhances gene mining efforts for industrial ene reductases by expanding enzyme panels with thermostable catalysts**. *ACS Catal* 2023, **4**:2576–2585.

39. * Joho Y, Vongsouthi V, *et al.*: **Ancestral sequence reconstruction identifies structural changes underlying the evolution of ideonella sakaiensis PETase and variants with improved stability and activity**. *Biochemistry* 2023, **62**:437–450.
ASR is explored to support and enhance the development of plastic-degrading (*i.e.* polyethylene terephthalate depolymerising) enzymes from serine hydrolase ancestors. The determinants for the substrate specificity of PETases are disclosed.

40. Kratzer JT, Lanaspa MA, Murphy MN, Cicerchi C, Graves CL, Tipton PA, Ortlund EA, Johnson RJ, Gaucher EA: **Evolutionary history and metabolic insights of ancient mammalian uricases**. *Proc Natl Acad Sci U S A* 2014, **111**:3763–3768.

41. Zakas PM, Brown HC, Knight K, Meeks SL, Spencer HT, Gaucher EA, Doering CB: **Enhancing the pharmaceutical properties of protein drugs by ancestral sequence reconstruction**. *Nat Biotechnol* 2017, **35**:35–37.

42. ** Thomas A, Evans BD, van der Giezen M, Harmer NJ: **Survivor bias drives overestimation of stability in reconstructed ancestral proteins**. *bioRxiv* 2022, https://doi.org/10.1101/2022.11.23.517659.
Harmer and co-workers investigate and rationalise the frequently introduced thermostability properties found in ancestrally reconstructed proteins.

43. Williams PD, Pollock DD, Blackburne BP, Goldstein RA: **Assessing the accuracy of ancestral protein reconstruction methods**. *PLoS Comput Biol* 2006, **2**:e69.

44. Hobbs JK, Shepherd C, *et al.*: **On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of Bacillus**. *Mol Biol Evol* 2012, **29**:825–835.

45. Trudeau DL, Kaltenbach M, Tawfik DS: **On the potential origins of the high stability of reconstructed ancestral proteins**. *Mol Biol Evol* 2016, **33**:2633–2641.

46. Wheeler LC, Lim SA, Marqusee S, Harms MJ: **The thermostability and specificity of ancient proteins**. *Curr Opin Struct Biol* 2016, **38**:37–43.

47. Saenko EL, Shima M, Sarafanov AG: **Role of activation of the coagulation factor VIII in interaction with vWf, phospholipid, and functioning within the factor xase complex**. *Trends Cardiovasc Med* 1999, **9**:185–192.

48. Schmidt AE, Bajaj SP: **Structure–function relationships in factor IX and factor IXa**. *Trends Cardiovasc Med* 2003, **13**:39–45.

49. * Knight KA, Coyle CW, Radford CE, *et al.*: **Identification of coagulation factor IX variants with enhanced activity through ancestral sequence reconstruction**. *Blood Adv* 2021, **5**:3333–3343.
Together with previous paper, this paper describes the discovery of biomedically promising variants of coagulations factors that were generated through ASR.

50. Simioni P, *et al.*: **X-linked thrombophilia with a mutant factor IX (factor IX Padua)**. *N Engl J Med* 2009, **361**:1671–1675.

51. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW: **Crystal structure of an ancient protein: evolution by conformational epistasis**. *Science* 2007, **317**:1544–1548.

52. Pillai AS, Chandler SA, Liu Y, *et al.*: **Origin of complexity in haemoglobin evolution**. *Nature* 2020, **581**:480–485.

53. * Hager M, Pöhler M, *et al.*: **Substrate affinity versus catalytic efficiency: ancestral sequence reconstruction of tRNA nucleotidyltransferases solves an enzyme puzzle**. *Mol Biol Evol* 2022, **39**. msac250.
The tRNA nucleotidyltransferases perform the essential addition of the CCA triplet to the 3′ terminus of all tRNAs. The paper shows that the extant enzymes avoid the reverse reaction, CCA hydrolysis, thanks to the weak affinity of the active site for their tRNA substrate, an interesting case of compensation between substrate affinity and catalytic activity.

54. Dobzhansky T: **Nothing in biology Makes Sense except in the light of evolution**. *Am Biol Teach* 1973, **35**:125–129.