Journal of Structural Biology 188 (2014) 213–224

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Amino acid sequence homologies in the hard keratins of birds and reptiles, and their implications for molecular structure and physical properties

R.D. Bruce Fraser [a,1], David A.D. Parry [a,b,*]

[a] Institute of Fundamental Sciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand
[b] Riddet Institute, Massey University, Private Bag 11-222, Palmerston North, New Zealand

ARTICLE INFO

Article history:
Received 22 July 2014
Received in revised form 24 October 2014
Accepted 28 October 2014
Available online 4 November 2014

Keywords:
β-Keratins
Feather
Scale and claw sequences
Archosaurs
Squamates
Rhynocephalians

ABSTRACT

Avian and reptilian epidermal appendages such as feathers, claws and scales exhibit a filament–matrix texture. Previous studies have established that both components reside within the same single-chain molecule. In the present study the homology in a wide range of aligned sequences is used to gain insights into the structure and function of the molecular segments associated with the filament and with the matrix. The notion that all molecules contain a β-rich 34-residue segment associated with the framework of the filament is reinforced by the present study. In addition, the residues involved in the polymerization of the molecules to form filaments are identified. In the Archosaurs (birds, crocodiles and turtles), and the Squamates (snakes and lizards) segments rich in glycine and tyrosine can be identified in the C-terminal domain. In Rhynocephalians (tuataras) and Squamates a similar segment is inserted at a specific point in the N-terminal domain. In some Archosaurian appendages (both avian and reptilian) segments rich in charged residues and cysteine are found in the N-terminal domain. The likely effect of these segments will be to soften the tissue without compromising its insolubility. The structure and role of the various molecular segments identified in this study and the way in which they might manifest themselves in terms of the physical properties of the particular epidermal appendage in which they appear are also discussed.

1. Introduction

All keratins exhibit a filament/matrix texture in cross-section when examined in the electron microscope (Rogers, 1959; Filshie and Rogers, 1962; Alexander, 1970; Maderson et al., 1972; Landmann, 1979; Alibardi et al., 2006, 2009). In mammalian keratins the intermediate filaments (8–10 nm diameter) contain a pair of proteins with a high proportion of α-helical coiled coil conformation flanked by short conserved homologous domains, while the matrix component consists of a multitude of proteins (keratin-associated proteins, KAPS), some of which are rich in cysteine (high-sulfur, HS and ultra-high sulfur, UHS) and others in glycine and tyrosine (HGT). The compositions and proportions of the matrix proteins vary from appendage to appendage and also with factors such as diet and health. In avian and reptilian hard keratins, however, the filament and matrix portions are combined within a single molecule consisting of a β-rich central domain, associated with the filament framework, and N-terminal and C-terminal domains thought to constitute the bulk of the matrix. It has been shown that the terminal domains contain both conserved and variables features (Fraser and Parry, 2011a). The overall diameter of the β-filament + matrix complex in avian and reptilian hard keratins is around 3.4 nm. It is also important to note, however, that there are now many examples where the intermediate filaments (IF) in mammalian hard corneous skin derivatives, such as hair, horns, claws and nails, co-exist with filaments closely akin to the β-filamentous structures characteristic of reptilian and avian hard keratins (see, for example, Dalla Valle et al., 2005, 2007a,b, 2009a,b, 2010; Hallahan et al., 2009). These β-keratins have been renamed "β-keratin-associated proteins" (β-KAPS) by Alibardi and colleagues to avoid confusion with the β-keratin derived by stretching α-keratin under conditions of pressure and temperature not experienced *in vivo*.

Abbreviations: HGT, high-glycine–tyrosine; HS, high-sulfur; UHS, ultra-high sulfur.

* Corresponding author at: Institute of Fundamental Sciences, Massey University, Private Bag 11-222, Palmerston North, New Zealand. Fax: +64 6 3505798.
E-mail address: d.parry@massey.ac.nz (D.A.D. Parry).
[1] Current address: 28 Satinay Drive, Noosa Parklands, Tewantin, Qld 4565, Australia.

http://dx.doi.org/10.1016/j.jsb.2014.10.012
1047-8477/

X-ray diffraction studies have yielded valuable information about the general form of the β-rich central domain (Astbury and Marwick, 1932; Bear and Rugo, 1951; Fraser et al., 1971) but insufficient data are available for the precise structure of the filament–matrix composite to be determined by conventional protein crystallographic methods. Following the determination of the amino acid sequence of emu feather (O'Donnell, 1973) infrared studies showed that a large enzymatic fragment from the interior of the molecule was rich in the β-sheet conformation (Suzuki, 1973). Later, a Fourier analysis of the distribution of the β-favoring and turn-favoring residues in the complete sequence (Fraser and MacRae, 1976) indicated that there was a central domain in which the chain looped backwards and forwards with a periodicity of eight residues resulting in an axial length of around 2.4 nm for this domain, a value equal to that of the spacing of a very prominent meridional reflection in the X-ray diffraction pattern of feather. A similar reflection in the X-ray patterns of a range of avian and reptilian hard keratins suggested that similar looped central domains might be present in all of these materials (Astbury and Marwick, 1932; Rudall, 1947; Fraser et al., 1972; Stewart, 1977). This was confirmed when the amino acid sequence of a reptilian keratin became available (Gillespie, 1990) and it was found that a close homology existed between the sequences of emu feather and goanna claw in this central domain (Fraser and Parry, 1996). Further extensive studies of reptilian keratins by Alibardi and coworkers (Alibardi et al., 2006, 2009; Dalla Valle et al., 2005, 2007a,b, 2009a,b, 2010; Toni et al., 2007), and avian keratins by Rogers and coworkers (Gregg et al., 1984; Presland et al., 1989; Whitbread et al., 1991) and by Greenwold and Sawyer (see, for example, Greenwold and Sawyer, 2010, 2013), showed that this homology is maintained over the entire range of hard β-keratins in birds and reptiles.

Modern day birds and reptiles (Fig. 1) can be traced back either to the Archosaurians (birds, crocodiles and turtles) or the Lepidosaurians, which include the Squamates (snakes and lizards) and the Rhynocephalians (tuataras). As most of the physical studies to date have been undertaken on avian keratins it was decided that the sequences of these keratins would be analyzed as a group before proceeding to the other Archosaurs and then finally to the Squamates and Rhynocephalians, in a search for features of similarity or difference. The rationale for so doing was the knowledge that the keratin proteins play a crucial role in specifying the physical properties of the appendage in which they occur. The aim of the present study was thus to learn more about the structure and role of the different molecular segments and how these features manifest themselves in the physical properties. This involves not only the molecular structure of keratin but also its ultrastructural topography. Determining how some of these properties arise in terms of the sequence characteristics of the keratin proteins involved has not previously been attempted but forms a key aim of the current research. It is, however, beyond the scope of the present study to relate the material properties of the different corneous material (beta-layer, alpha-layer, feather, claw, scale) to specific proteins, though this remains a longer-term aim.

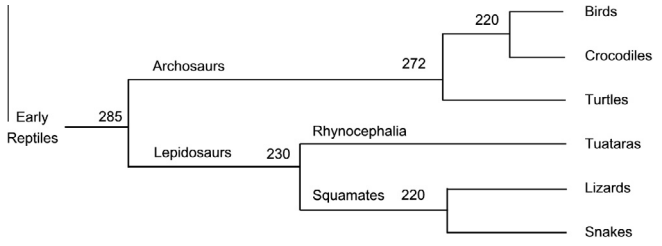

Fig.1. The phylogeny of modern birds and reptiles. The estimates of the branching (in millions of years) are based on studies of the mitochondrial DNA (Rest et al., 2003).

2. Experimental

Sequence data covering examples of both the avia and reptilia and of different epidermal appendages (feathers, claws, scales, keratinocyte) were collected from the NCBI Reference Sequence Database and from the Protein Knowledge Base (UniProKB). Details are available in a file in FASTA format in the Supplementary Material. The sequence databases, however, have rarely included any information on the physical origin of the sequence (scale, claw, etc), other than those relating to feather. We have been loath to make an assumption based on expression profiles (not always available) as to the particular appendage involved when the authors responsible for determining the sequences have failed to make the correlation themselves.

Any selection from the multitude of available sequences is of necessity arbitrary but the guiding principles were to maximize the range of species, appendages, character and size. Particular attention was given to the inclusion of unusual sequences with the aim of learning more about the minimum requirement for filament formation and the composition of the matrix. The selection was made without reference to gene data. Multiple alignments of both the entire data set and selected groups were carried out using Unipro Ugene as a front end for various alignment schemes. The alignments differed somewhat according to both the scheme and to the values allotted to the various parameters. The choice of alignment scheme is to a certain degree subjective but the one adopted in this work was the MUSCLE program (EMBL-EBI) using the default parameters. The results obtained with Clustal W (EMBL-EBI) were similar but not identical. The alignments were displayed using JalView (The Barton Group, University of Dundee, Scotland, UK).

3. Results and discussion

Most of the sequences used have been derived from nucleic acid data and, as a consequence, direct experimental evidence for participation of these chains in filamentous structures is lacking. However, on the basis of sequence homology with those relatively few chains derived directly by protein extraction and subsequent sequencing using protein chemical methods it has been assumed in this work that the great majority of these chains do indeed self-assemble into a filamentous structure, and are therefore appropriate for direct comparison. However, it remains a possibility that some sequences used in these analyses contain key mutations that prevent their involvement in filament formation. These chains would then, presumably, have a different but no less important role *in vivo*.

A preliminary alignment of the entire data set (Fig. 2) showed that there are highly homologous segments in the bulk of the chosen sequences. Separate alignments were then carried out for avian Archosaurs, reptilian Archosaurs, and Lepidosaurs. Visual inspection of the homology in Fig. 2 reveals that certain segments persist in a slightly modified form in all avian and reptilian keratins whilst other segments vary greatly in length but have certain characteristics in common.

In our earlier studies, based partly on amino acid composition, it was suggested that the sequences of avian and reptilian keratins could be divided into five distinct segments according to the scheme shown in Fig. 3 (after Fraser and Parry, 2011a). In that work three distinct domains were recognized (a) a highly homologous central domain around 34 residues in length, containing a high proportion of β-favoring residues, and (b) N-terminal and