Volume 333, number 1,2, 99–102 FEBS 13144 October 1993

Structural relationship of streptavidin to the calycin protein superfamily

Darren R. Flower*

Department of Physical Chemistry, Fisons Plc, Pharmaceuticals Division, R & D Laboratories, Bakewell Rd, Loughborough, Leicestershire, LE11 ORH, UK

Received 29 July 1993; revised version received 2 September 1993

Streptavidin is a binding protein, from the bacteria *Streptomyces avidinii*, with remarkable affinity for the vitamin biotin. The lipocalins and the fatty acid-binding proteins (FABPs), are two other protein families which also act by binding small hydrophobic molecules. Within a similar overall folding pattern (a β-barrel with a repeated +1 topology), large parts of the lipocalin, FABP, and streptavidin molecules can be structurally equivalenced. The first structurally conserved region within the three-dimensional alignment, or common core, characteristic of the three groups corresponds to an unusual structural feature (a short 3_{10} helix leading into a β-strand, the first of the barrel), conserved in both its conformation and its location within their folds, which also displays characteristic sequence conservation. These similarities of structure and sequence suggest that all three families form part of a larger group: the calycin structural superfamily.

Streptavidin; Calycin; Lipocalin; Fatty acid-binding protein; Protein structure comparison; Structural superfamily

1. INTRODUCTION

Streptavidin is a small soluble protein isolated from the bacteria *Streptomyces avidinii*, which shares with its hen egg-white homologue avidin, a remarkable affinity for the vitamin biotin [1]. These proteins have found a range of important applications in biotechnology. The structure of streptavidin has been reported, to high resolution, by two independent studies. Weber et al. [2] report the structure of apo-streptavidin to 1.8 Å resolution and its biotin complex to 2.6 Å. Hendrickson et al. [3] report the structure of a truncated streptavidin to 2.0 Å. As an aside to their description of its structure, an eight-stranded antiparallel up-and-down barrel β-barrel, Hendrickson et al. comment on the similarity of streptavidin's topology to that of the lipocalins. In reviewing this proposal, Cowan et al. [4] refute the suggestion that these groups of proteins are related. The issue of their similarity remains unresolved.

A recent study has shown that the lipocalins and a related family, the fatty acid-binding proteins (FABPs), both of which function by binding small hydrophobic molecules, form an overall structural superfamily, the calycins [5], characterized by a conserved folding pattern, within which large parts of their structures can be superimposed closely in three-dimensions, and by a common N-terminal sequence motif. This paper reports a study showing that streptavidin, and by inference other avidins, is also a member of the calycin superfamily, related, quantitatively and qualitatively, to both the lipocalins and FABPs in much the same way that they are related to each other.

*Corresponding author. Fax: (44) (509) 611 222.

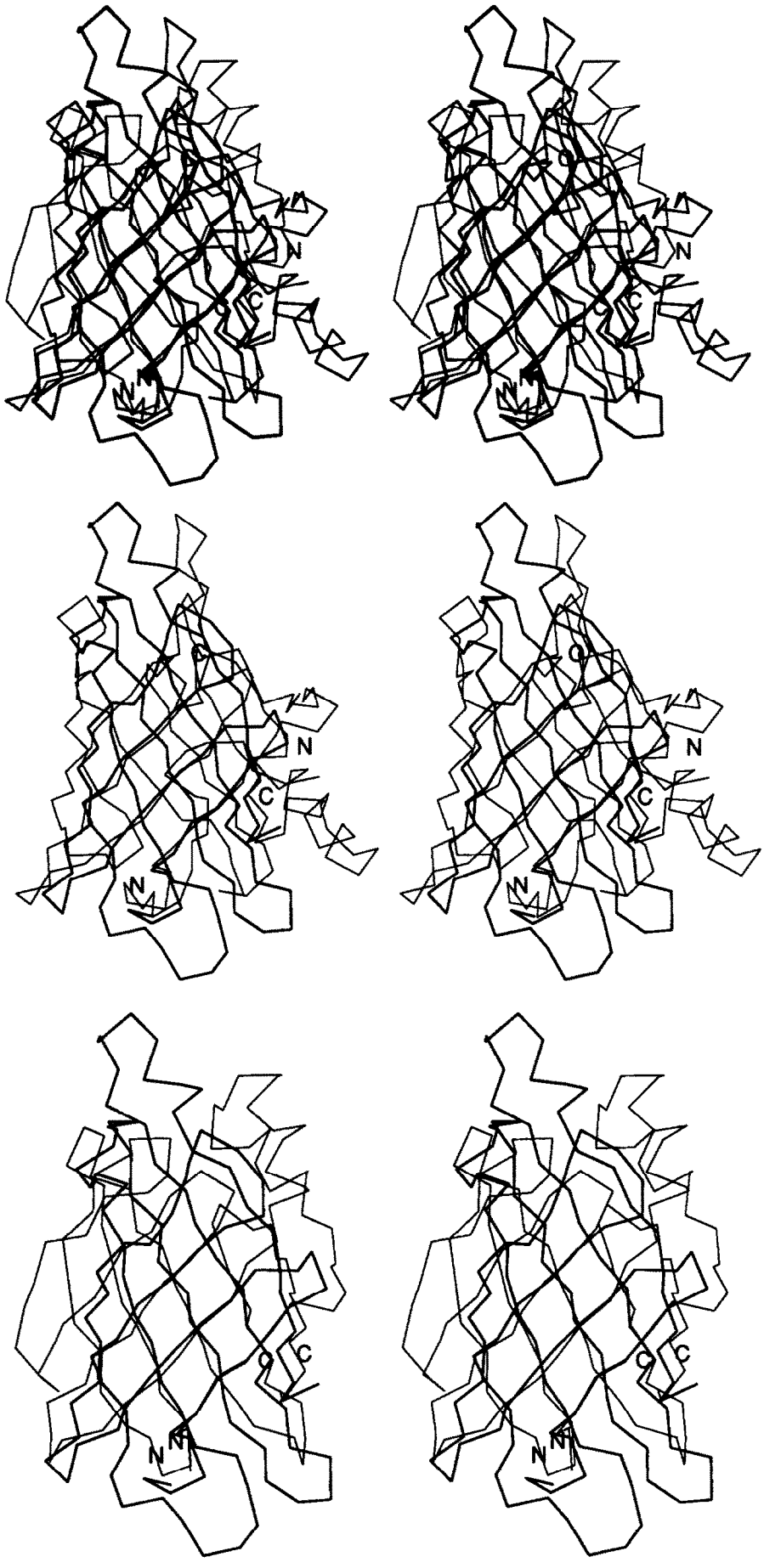
Abbreviations: FABP, fatty acid-binding proteins; I-FABP, rat intestinal fatty acid-binding protein; RBP, human retinol binding protein; RMS, root mean squared; SCR, structurally conserved region

2. MATERIALS AND METHODS

Coordinates for the structures of human retinol binding protein (databank code 1RBP) [4], rat intestinal fatty acid binding protein (code 2IFB) [6] and streptavidin (code 1STP) [2] were obtained from the Brookhaven Protein Databank [7]. Structures were compared manually by the interactive manipulation of C_α coordinates using the molecular graphics program WHAT-IF [8]. Residues were deemed equivalent if their superimposed C_α coordinates deviated by less than 3.0 Å.

3. RESULTS

The structures of a representative lipocalin, retinol-binding protein (RBP) [4], and a representative FABP, rat intestinal fatty acid-binding protein (I-FABP) [6], were superimposed manually onto that of streptavidin [5], using interactive computer graphics, in order to facilitate their quantitative structural comparison. From this the equivalent parts of their structures were established allowing a 'common core' characteristic of the underlying fold to be determined. This multiple superposition is shown graphically as a set of superposed structures in Fig. 1, while a sequence alignment based on it is given in Fig. 2. The three structures can be equivalenced so that the initial four and final two strands of each barrel match to form contiguous structurally conserved regions (SCRs), with little or no correspondence between strands from the centre of their closed β-sheets or between any intervening loop regions. The RMS deviation between equivalenced residues in


a
N
C
N
C
b
N
C
N
C
c
C
N
C
N

```
RBP       ERDCRVSSFRVKENFDKARFSGTWYAMAKKDPEG-------------LFLQDNIVAEFSVDETGQMSAT
                           *********                                *******     *****
I-FABP                    AFDGTWKVDRNENYEKFMEKMGINVVKRKLGAHDNLKLTITQEG-NKFTVK
                          *********                                *******     *****
STP                     AEAGITGTWYNQL-----------------------GSTFIVTAGADGALTGT

RBP       AKGRVRLLNNWDVCADMVGTFTDTEDPAKFKMKYWGVAS-------------FLQKGNDDHWIVDTDYD
          ***               *****
I-FABP    ESSNF---------RNIDVVFELGVDFAYSLADGTELTGTWTMEGNKLVGKFKRVDNGKELIAVREISG
          ***               *****
STP       YESAVG-----NAESRYVLTGRYDSAPATDGSGTALGWTVAWK-----NNYRNAHSATTWSGQYVGGAE

RBP       TYAVQYSCRLLNLDGTC----ADSYSFVFSRDPNGLPPEAQKIVRQRLCLARQYRLIVHNGYC
          ******                          *****
I-FABP    NELIQTYTYEGV-----------EAKRIFKKE
          ******                          *****
STP       ARINTQWLLTSGTTEANAWKSTLVGHDTFTKV
```

Fig. 2. Sequence alignment derived from structural superimpositions. The sequences of streptavidin (STP), a lipocalin (human retinol binding protein; RBP), and a FABP (rat intestinal fatty acid-binding protein; I-FABP) were aligned on the basis of equivalences determined from a multiple superimposition of their three-dimensional structures. The symbol (*) between residues indicates an equivalence when an optimal fit in three dimensions has been made. Aligned residues without a star are not structurally equivalent and do not form part of the common core.

streptavidin and RBP is 1.86 Å, between streptavidin and I-FABP is 1.62 Å, and between RBP and I-FABP is 1.59 Å.

The structure of streptavidin is dominated by eight β-strands (labelled A to H) which form an antiparallel β-barrel enclosing an internal ligand (biotin) binding site. These eight strands are linked by a succession of seven +1 connections (labelled to L1 to L7); the simplest possible topology for a closed sheet. The seven loops, although they vary in size and conformation, are all typical of short β-hairpins. A short 3_{10} helix, directly before strand A, helps to close off one end of the barrel. The arrangement of these features within the fold is shown schematically in Fig. 3A. The topology of all three structures is very similar: each form antiparallel β-barrels with repeated +1 connections (see Fig. 3), although the FABP barrel is more flattened or elliptical than either that of the lipocalins or streptavidin and is not continuously hydrogen bonded, with a wide discontinuity between strands D and E. Although apparently more similar, the avidin barrel is both more circular and compact, in cross-section, than that of the lipocalins. It is this difference in size and shape which explains the inability to superimpose more of their structures. In the lipocalins and FABPs, loop L1 is a large and atypical Ω-loop which forms an important feature of their structures capping the internal binding site, while in streptavidin loop L1 is, by contrast, a short, typical β-hairpin.

Within a similar topological pattern, much of the lipocalin, FABP, and streptavidin molecules can be structurally equivalenced: this is indicative of a close relationship between their characteristic folds. Moreover, inspection of the alignment in Fig. 2 indicates that the first SCR of the common core also displays discernable sequence similarity (8 pairwise identities), including the complete preservation of key residues (the GTW triplet), while the other SCRs (with between 1 and 3 pairwise identities), and indeed the sequences as a whole, do not show such sequence conservation. This characteristic sequence motif, which corresponds to an unusual structural feature (a short 3_{10} helix leading into a β-strand, the first of the barrel), conserved in both its conformation and its location within the fold of the lipocalins, FABPs, and streptavidin, has been shown to be present consistently in all members of the calycin superfamily [5].

4. DISCUSSION

The structural analysis presented here demonstrates that within a similar folding pattern (an antiparallel

←

Fig. 1. Superimposition of three-dimensional structures. The superimposed structures of streptavidin, RBP, and I-FABP are shown as a series of three stereo pairs of C_α traces. N and C termini are marked. (a) Three-way superimposition of RBP (thin lines), streptavidin ((medium thick lines), and I-FABP (thick lines). (b) Two-way superimposition of RBP (thin lines) and streptavidin (thick lines). (c) Two-way superimposition of I-FABP (thin lines) and streptavidin (thick lines).

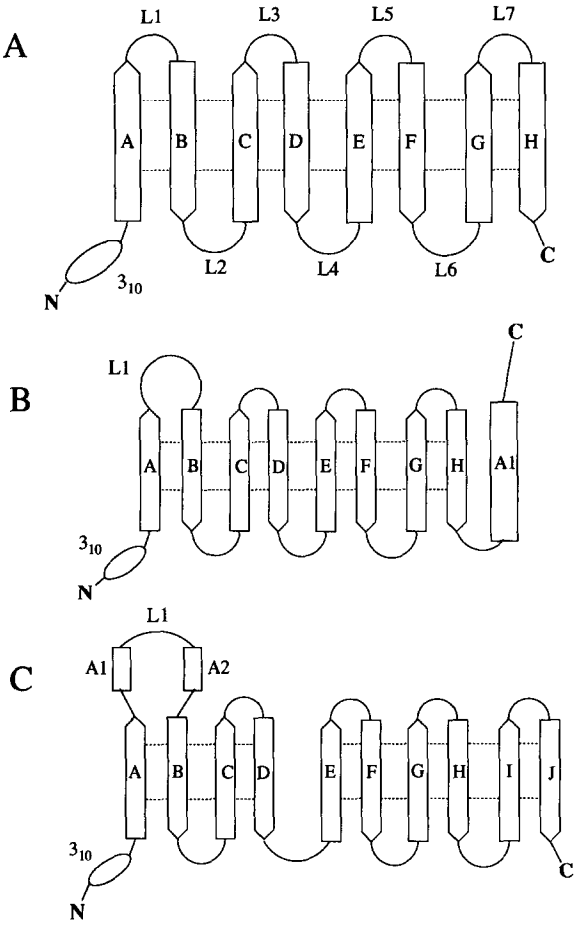


Fig. 3. Schematic drawings of characteristic structural features. (A) The avidin fold. The eight β-strands of the streptavidin barrel are shown as arrows and labelled A–H. Linking β-hairpin loops are labelled L1–L7. The conserved N-terminal 3_{10} helix is marked. (B) The lipocalin fold. The eight-strands of the barrel are shown as arrows and labelled A–H. The conserved C-terminal α-helix (labelled A1) is also marked, as is large linking loop L1. (C) The FABP fold. The 10-strands are shown as arrows and labelled A–J. Short conserved α-helices within loop L1 are labelled A1 and A2. In each diagram the hydrogen bonded connection of two strands is indicated by a pair of dotted lines between them.

β-barrel, with a repeated +1 topology, possessed of an internal ligand binding site), large amounts of the lipocalin, FABP, and streptavidin structures can be equivalenced to form a number of significant discrete SCRs. The comparison of hen egg avidin and streptavidin recently reported by Pugliese et al. [9], which showed them to have only trivial structural differences, suggests that the structural similarity reported here for streptavidin should hold for all avidins. Equally, comparisons of known FABP [10] and lipocalin [5] structures indicates that this relationship is generally true across all of the families. These results indicate that, notwithstanding the very low global sequence similarity between any of these three groups, a close structural relationship exists between these protein families. In turn, this suggests that the avidins form part of the calycin structural superfamily, a group previously reported to comprise the lipocalins and FABPs [5].

It is well known that protein structure is better conserved in evolution than protein sequence and that in extreme cases close relationships are not apparent at the sequence level: two contrasting views of this phenomena are emerging. One holds that despite retention of close structural propinquity, divergence from some common ancestor has occurred beyond a point at which any significant sequence similarity remains [11]. The other predicates a model whereby structures resemble each other, more or less, by chance because of the limited number of thermodynamically stable possible folds [12]. With regard to the calycins, a further example of this phenomenon, the observation that all members of the superfamily retain a common, characteristic, and conserved N-terminal sequence motif argues for the former explanation of their obvious similarities and supports the conjecture that the lipocalins, FABPs, and streptavidin share an evolutionary relationship with their differences reflecting functional specialization.

REFERENCES

[1] Richards, F.M. (1990) Methods Enzymol. 184, 3–5.
[2] Weber, P.C., Ohlendorf, D.H., Wendoloski, J.J. and Salemme, F.R. (1989) Science 243, 85–88.
[3] Hendrickson, W.A., Pahler, A., Smith, J.L., Satow, Y., Merritt, E.A. and Phizackerley, R.P. (1989) Proc. Natl. Acad. Sci. USA 86, 2190–2194.
[4] Cowan, S.W., Newcommer, M.E. and Jones, T.A. (1990) Proteins: Structure Function and Genetics 8, 44–61.
[5] Flower, D.R., North, A.C.T. and Attwood, T.K. (1993) Prot. Sci. 2, 753–761.
[6] Sacchetini, J.C., Gordon, J.L. and Banaszak, L.J. (1989) J. Mol. Biol. 208, 327–345.
[7] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[8] Vriend, G. (1990) J. Mol. Graphics 8, 52–56.
[9] Pugliese, L., Coda, A., Malcovati, M. and Bolognesi, M. (1993) J. Mol. Biol. 231, 698–710.
[10] Zanotti, G., Scapin, S., Spadon, P., Veerkamp, J.H. and Sacchettini, J.C. (1992) J. Biol. Chem. 267, 18541–18550.
[11] Murzin, A.G. and Chothia, C. (1992) Curr. Opin. Struct. Biol. 2, 895–903.
[12] Finkelstein, A.V., Gutun, A.M. and Badretdinov, A.Y. (1993) FEBS Lett. 325, 23–28.