Vol. 180, No. 1, 1991
October 15, 1991

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
Pages 69-74

Mouse oncogene protein 24p3 is a member of the Lipocalin protein family

D.R.Flower, A.C.T.North, and T.K.Attwood

Department of Biochemistry and Molecular Biology
University of Leeds
Leeds, UK
LS2 9JT

Received August 8, 1991

Summary: Rigorous new methods of protein sequence analysis have been applied to the lipocalins, a diverse family of ligand binding proteins. Using three conserved sequence motifs to search for similar patterns in a large sequence database, the size and composition of this protein family have been defined in an automatic and objective way. It has allowed the identification of an existing sequence, mouse 24p3 protein, as a lipocalin and the possible rejection of other putative members from this protein family. On the basis of this newly discovered homology, a possible function for mouse 24p3 protein is proposed.

In recent years, the complementary disciplines of protein crystallography and protein sequence analysis have identified a new protein superfamily based on the similarity of both their primary and tertiary structures (1,2). This family, the lipocalins, is composed mainly of extracellular ligand binding proteins displaying high specificity for small hydrophobic molecules (3). Representative members of this family include plasma retinol binding protein, β-lactoglobulin, and major urinary protein. The family is constantly expanding as new proteins are sequenced and classified from their sequence homology as lipocalins. Such examples include: Ch21 protein (4,5), aphrodisin (6), Von Ebner's Gland protein (7), and probasin (8). Moreover, a number of studies have extended the family by identifying extant sequences as lipocalins by homology searching of sequence databases. Examples include α-1-acid glycoprotein (9), human $C8_\gamma$ (10), rat epididymal secretory protein (11), and prostaglandin D synthase (12). In

Abbreviations: **RBP**, retinol binding protein; β-**lg**, β-lactoglobulin; **α1MG**, α-1-microglobulin; **MUP**, major urinary protein; **ADSP**, Algorithms and Data Structures for Protein sequence analysis; **SOMAP**, Screen Orientated Mutiple Alignment Procedure.

0006-291X/91 $1.50

this paper we describe the application of a powerful new method of sequence analysis to this family. This approach, embodied in the ADSP software system (13), can define and then refine the composition of protein families on the basis of their sequence homology, even when this is restricted to only a few short conserved motifs.

MATERIALS AND METHODS

By using the interactive multiple alignment program SOMAP (14), an initial sequence alignment was constructed which contained a small set of lipocalin sequences, representative of those proteins from which the family was originally defined. From this alignment a set of conserved sequence motifs was extracted and used to scan the OWL database (15), version 11.0, for similar motifs, using the systematic procedure embodied in the ADSP system (13,16). The first such scan identified further sequences which shared all of these conserved features; revised motif sets were extracted from this new group of sequences and used to rescan the database. This iterative process was continued to convergence, ie until the identified sequence sets remained constant between scans. Comparison of the final set, generated by this blind automatic procedure, with sequences classified in the literature as lipocalins allows a more rigorous appraisal of the constitution of the family.

RESULTS

From the initial alignment it is again clear that the global sequence homology of the lipocalin family is unusually low, despite their very similar three dimensional structures (2); however, within this low overall similarity there are three short motifs which are highly conserved between members of the family (Fig 1). These three features, although well separated in the sequence, occur very close together in the three dimensional structure of the folded protein. This intimate juxtaposition is conserved in all the known lipocalin crystal structures, and has been noted before for two of the three sequence features described here, when it was suggested that they might constitute a receptor binding site (17). After convergence of the scanning procedure using the three motifs, ADSP had identified the majority of sequences known from our surveys of the literature, but had also identified one further member, mouse 24p3 protein (18), not previously known to be a lipocalin. A number of sequences classified by others as lipocalins: aphrodisin (6), Von Ebner's Gland protein (7), probasin (8), rat odorant binding protein (19), and α-1-acid glycoprotein (9), did not form part of this final set, as they do not possess all of the three conserved motifs. Figure 2 shows motif discrimination profiles for bovine β-lg (a sequence from the initial alignment), the new member 24p3 protein, and a supposed member aphrodisin. Comparison of these plots

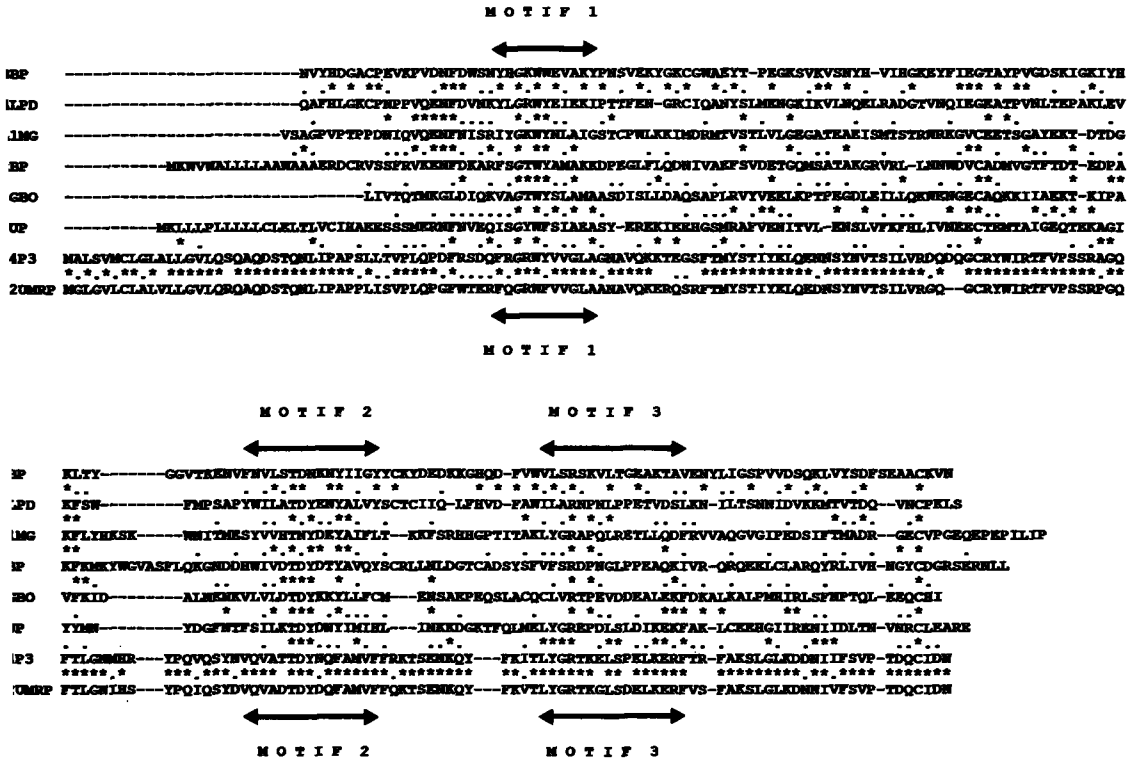

Figure 1. A multiple alignment of lipocalin family members. The first six sequences are in order, **BBP:** Bilin-binding protein from pieris brassicae, **ALPD:** human apo-lipoprotein D, **A1MG:** human α-1-microglobulin, **RBP:** human plasma RBP, **LGBO:** bovine β-lg, and **MUP:** mouse major urinary protein. These sequences formed the starting alignment from which the three initial conserved motifs were extracted. These motifs are marked. The last two sequences are **A2UMRP:** rat α-1-Microglobulin-related-protein and **24P3:** mouse 24p3 protein, the alignment shows the high degree of identity between them and the conservation of the three characteristic motifs evident in the larger group.

demonstrates clearly how the motif sets act together as a powerful diagnostic of family membership: members give obvious hits for all motifs.

DISCUSSION

Using the ADSP method, it has been possible to define the composition of the lipocalin protein family in an automatic and objective way. Thus we may have considerable confidence in the high degree of relatedness between the members of this set. It is possible to augment the set of core members with a number of other proteins which have been classified in the literature as lipocalins on the basis of limited global homology to one or more lipocalin members, but which do not share all the motifs fully

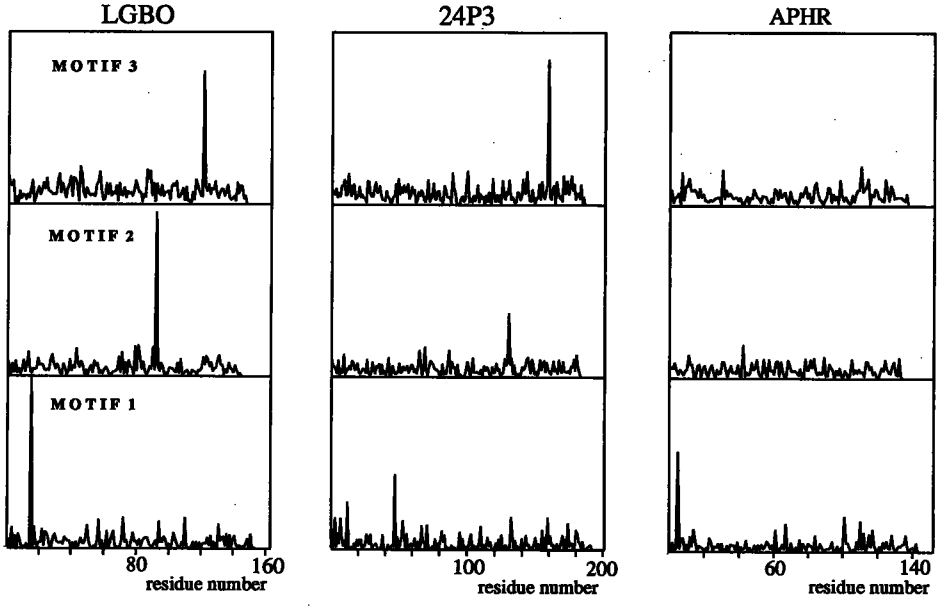

Figure 2. Motif discrimination plots. Discriminator plots for three sequences are shown: **LGBO** Bovine β-lg, a sequence from the initial alignment, **24P3** mouse 24p3 protein, the newly discovered member of the family, and **APHR** aphrodisin, thought previously to be a lipocalin. These plots indicate the quality and location of the matching of a motif with a protein sequence; the Y ordinate for each motif representing percentage matching of the sequence and motif. Clearly all motifs match well for β-lg and 24p3, whilst only the first matches for Aphrodisin.

consistent within the core group. These proteins may represent more highly diverged relatives with only limited structural or functional similarity to other members of the lipocalin family.

Mouse 24p3 protein is the product of a single gene, 24p3, and is notable because of the 7-10 fold increase in its expression in cultured mouse kidney cells infected by SV40 or other viruses (18). The authors stated that it had no homology to any known sequence and presented no evidence for the function of the mature protein. Our results give insight into both of these aspects. Its membership of the lipocalin family, deduced here from motif matching, is corroborated both by its size and pattern of cysteine residues, which are consistent with that observed for other family members, and by its high global homology with certain other previously identified members, especially rat α-2-microglobulin-related-protein (20). We find these two sequences to be 76% identical with 1 short insertion; both then show the next greatest similarity to mouse

major urinary proteins and rat α-2-microglobulins (see Figure 1). Mouse 24p3 protein and rat α-2-microglobulin-related-protein may represent the same or very similar proteins derived from different species; the sequence divergence between species for other lipocalin members, such as β-lactoglobulin, is of a comparable magnitude.

Although the precise functions of mouse 24p3 protein and rat α-2-microglobulin-related-protein have yet to be determined, it is known that a number of other lipocalins bind retinoids, such as retinol (RBP, β-lg, α1MG (3)) and retinoic acid (rat epididymal secretory protein (11)): it is well known that both these molecules act in the potentiation and control of cell regulation, differentation, and mitogenesis. Moreover, another lipocalin, purpurin (21), which also binds retinol, is believed to function in the control of cell differentiation and survival. Ch21 protein (4,5) has also been implicated in cell regulation: its expression is suppressed during cell division, although the ligand and mode of action of this protein are as yet unknown. Thus mouse 24p3 protein, and by analogy rat α-2-microglobulin-related-protein, may function in the control of cell regulation, in normal and/or transformed cells, through the transport of retinol, retinoic acid, or other lipophilic molecules. The lipocalin family may yet prove to be even more widespread and important than has been thought hitherto, with crucial roles in the most fundamental of biological processes.

Acknowledgments

D.R.Flower is the recipient of a Wellcome Prize studentship. We thank the SERC for the provision of computing resources.

REFERENCES

1. Pervais, S. and Brew, K. (1985), *Science* **228**, 335-337.
2. North, A. C. T. (1991), *Biochem. Soc. Symp.* **57**, 35-48.
3. Godavac-Zimmermann, J. (1988), *Trends in Biochem. Sci.* **13**, 64-66.
4. Cancedda, F.D., Dozin, B., Rossi, F., Molina, F., Cancedda, R., Negri, A., and Ronchi, S. (1990), *J.Biol.Chem.* **265**, 19060-19064.
5. Bedard, P-A., Yannoni, Y., Simmons, D.L., and Erikson, R.L. (1989), *Mol. Cell. Biol.* **9**, 1371-1375.
6. Henzel., W.J., Rodriguez, H., Singer, A. G., Stults, J.T., Macrides, F., Agosta, W.C., and Nial, H. (1988), *J.Biol.Chem.* **263**, 16682-16687.
7. Schmale, H., Holtgreve-Grez, H., and Christiansen, H. (1990), *Nature* **343**, 366-369.
8. Spence, A.M., Sheppard, P.C., Davie, J.R., Matuo, Y., Nishi, N., McKeehan, W.L., Dodd, J.G., and Matusik, R.J. (1989), *Proc. Nat. Acad. Sci. USA* **86**, 7843-7847.

9. Pervais, S., and Brew, K. (1987), *FASEB J.* **1**, 209-214.
10. Hunt, L.T., Elzanowski, A., and Barker, W.C. (1987), *Biochem. Biophys. Research Communications* **149**, 282-288.
11. Brooks, D.E., (1987), *Biochem. Int.* **14**, 235-240.
12. Nagata, A., Suzuki, Y., Igarashi, M., Eguchi, N., Toh, H., Urade, Y., Hayaishi, O. (1991), *Proc. Natl. Acad. Sci.* **88**, 4020-4024.
13. Parry-Smith, D.J., and Attwood, T.K., (1991), *J.Mol.Biol.*, submitted.
14. Parry-Smith, D.J., and Attwood, T.K., (1991), *CABIOS* **7**, 233-235.
15. Bleasby, A.J., and Wootton, J.C. (1990), *Prot. Eng.* **3**, 153-159.
16. Attwood, T.K., and Findlay, J.B.C., (1991), *J.Mol.Biol.*, submitted.
17. North, A. C. T. (1989), *Int. J. Biol. Macromol.* **11**, 56-58.
18. Hraba-Renevey, S., Turler, H., Kress., H., Salomon, C., and Weil, R. (1989), *Oncogene* **4**, 601-608.
19. Pevsner, J., Reed. R.R., Feinstein, P.G., and Snyder, S.H. (1988), *Science* **241**, 336-339.
20. Chan, Y-L., Paz, V., and Wool, I.G. (1988), *Nucleic Acids Research* **16**, 11368.
21. Berman, P., Gray, P., Chen, E., Keyser, K., Ehrlich, D., Karten, H., Lacorbiere, M., Esch, F., and Schubert, D. (1987), *Cell* **51**, 135-142.