Vol. 155, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1988 Pages 32-37

STRUCTURE OF THE LYSOSOMAL SPHINGOLIPID ACTIVATOR PROTEIN 1 BY HOMOLOGY WITH INFLUENZA VIRUS NEURAMINIDASE

Michel Potier

Service de Génétique Médicale, Hôpital Sainte-Justine,
Université de Montréal, Montréal, Québec, Canada H3T 1C5

Received July 12, 1988

The sphingolipid activator protein 1 (SAP-1) increases the rate of hydrolysis of sphingolipids in the lysosome by apparently bringing together the substrate and the corresponding hydrolytic enzyme. This implies specific recognition of both the substrate and enzyme by SAP-1. However, binding domains in SAP-1 and recognition mechanisms involved are unknown. Amino acid sequence comparison of SAP-1 with influenza virus neuraminidase (EC 3.2.1.18, FLU NA) indicates that functional amino acid residues in or near the sialic acid binding site of FLU NA are also found at equivalent positions in the first 48 N-terminal amino acids of SAP-1. This region of homology allows to propose folding of the SAP-1 polypeptide chain by comparison with known crystallographic structure of FLU NA and identify a potential domain for lysosomal enzyme recognition through sialic acid binding. There is also a region of 10 amino acid residues near the C-terminal end of SAP-1 which has a strong propensity to form an α-helix with amphiphilic properties of lipid-binding helices. This domain in SAP-1 is probably responsible for the lipid(substrate)-binding function of SAP-1.

The degradation of sphingolipids in the lysosome is carried out by specific hydrolytic enzymes with the help of small molecular weight activator proteins (1-4). The function of these activator proteins is apparently to help lipid substrate cleavage by bringing together substrate and hydrolytic enzyme (5). The sphingolipid activator protein 1 (SAP-1) is a homodimer (subunit M_r 8 000 to 11 000, depending upon glycosylation (6)) which increases the rate of hydrolysis of sulfatides, GM1-ganglioside and trihexosylceramide substrates by the corresponding lysosomal enzymes arylsulfatase A, β-galactosidase and α-galactosidase (5,7,8).

Although the entire amino acid sequence of SAP-1 is known (67 residues, ref. 9), the functional and structural domains implicated in specific binding to the substrate and enzyme have not been identified and the 3D structure of SAP-1 is unknown. A method to detect domains in proteins is by comparison with proteins with known structure and functions.

0006-291X/88 $1.50

Many structural motifs have been conserved during evolution and they are found as structural and functional domains in a variety of different proteins (10).

In a search for specific recognition mechanisms between SAP-1 and the lysosomal enzymes, I considered the possibility of a sialic acid-mediated recognition mechanism (11) and compared SAP-1 amino acid sequence with that of influenza virus (FLU) neuraminidase (EC 3.2.1.18, NA) as a model of a sialic acid binding protein. Considerable data on inter-strain variations of FLU NA amino acid sequence have been obtained in the literature (12) and many conserved polar amino acid residues potentially implicated in sialic acid binding and substrate hydrolysis have been identified (13). Some have been localized in and around the active site pocket of FLU NA by X-ray crystallography (13,14). I report here that a region of SAP-1 including the first 48 N-terminal amino acids contains several of these active-site polar-residues of FLU NA. I propose a folding structure of SAP-1 based on its sequence homology with NA and on predicted secondary structure from the amino acid sequence. Finally, the model defines two structural and functional domains in SAP-1: one for enzyme binding and one for sphingolipid binding.

METHODS

Influenza virus neuraminidase sequences were taken from the Genbank database. Research for similarity with other protein sequences was carried out in the Protein Sequence Database of the Protein Identification Resource (release 12.0) using the FASTP software (15).

Secondary structure of SAP-1 in its C-terminal region was predicted according to Garnier et al. (16). Mean hydrophobicity, <H>, and mean hydrophobic moment per residue, $<\mu_H>$, of the predicted C-terminal α-helix of SAP-1 was computed according to Eisenberg et al. (17).

RESULTS AND DISCUSSION

Evolution conserves functional amino acids, secondary and tertiary structures of proteins rather than amino acid sequence. The inter-strain variability of FLU NA sequences identify many conserved amino acid residues which were localized near or in the active site pocket of FLU NA by X-ray crystallographic analysis (13). Among the residues that may be directly in contact with the sialic acid of the substrate and may be implicated in sialic acid recognition are the His 274, Glu 276 and Glu 277 and the amino acids around these residues: Asp 243 and Arg 292 (replaced by Lys 292 in some FLU NA strains) (Fig. 1). I report that SAP-1 contains the same polar amino acid residues at equivalent positions as in FLU NA (Fig. 1). In addition, I found four invariant Cys residues in SAP-1 corresponding to positions

```
 ß3S1   ß3S2            ß3S3      ß3S4             ß4S1    ß4S2

-----  --------       ---------- --------         ------- --------
   +    +     *                              *  **+             +*
ACNCIGGDCYLMITDGSASGISKCRFLKIREGRIIKEILPTSRVEHT EECTCGFASNKTIECACR  B N
ECTCVNGSCFTIMTDGPSDGLASYKIFKIEKGKIVHISPLSGSNSHY EECSCYPDTGKV MCVCR  A N1
ECVCINGTCTVVMTDGSASGRADTKILFIEEGKIVHISPLSGSAQHV EECSCYPRYPGV RCICR  A N2
GDVC  QDCIQMVTDI         QTAVRTNSTFVQAL    VEHVKEECDRLG  PGM ADICK  SAP1
----    ------           ----- !  ----            ----     -------

NYISQYSEIAIQMMMHMQP                                                 SAP1

       ----------
         helix
```

Fig. 1: Amino acid homologies between SAP-1 and FLU NA subtypes. The alignments of FLU NA type B (Lee/40, residues 228 to 292), FLU NA N1 (WS/33, residues 214 to 277) and N2 (Victoria/75, residues 229 to 292) are from ref. 12. The SAP-1 sequence is from ref. 9. The plus sign above the amino acid sequences indicates conserved Cys residues. The asterisks denote those charged amino acid residues that have been described by Colman et al. (13) to be close to the sialic acid binding site. The lines above the amino acid sequences indicate the regions described as β-sheets in the FLU NA structure. The lines below the SAP-1 sequence show the proposed location of β-sheets and an α-helix in SAP-1. The ! denotes the glycosylation site. The nomenclature of β-sheets and amino acid numbering is that in ref. 14.

of disulfide bridges Cys 232-Cys 237 and Cys 278-Cys 291 in FLU NA which each links two anti-parallel β-sheets (Fig. 1 and 2). Although the homology between SAP-1 and FLU NA was significant by structural and functional criteria, it was not detected by the FASTP software searching the PIR database of protein sequences. In fact, the similarity is not statistically significant, the score being lower than two standard deviations of the randomized mean score (15).

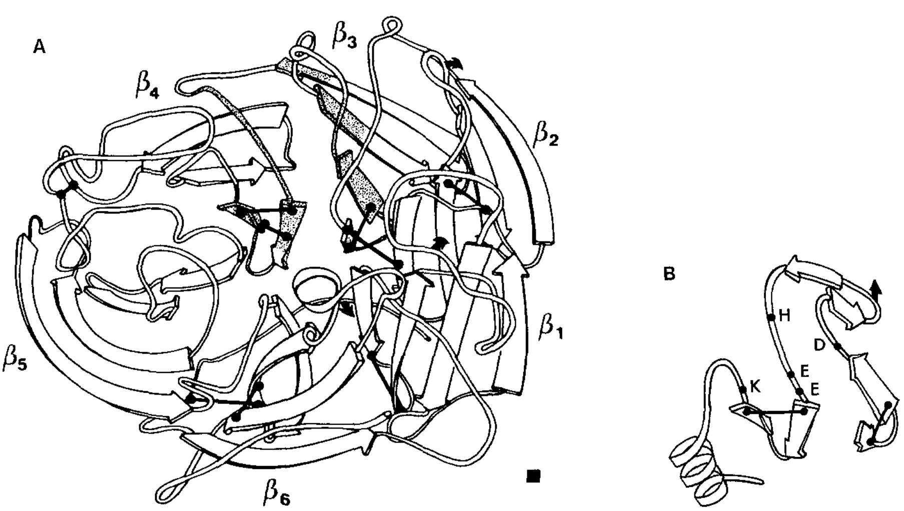

Fig.2: (A) Folding of influenza neuraminidase N2 subunit taken from ref. 14. The regions of homology with SAP-1 are darkened. (B) Model of SAP-1 folding based on homology with neuraminidase and prediction of a C-terminal helix according to Garnier et al. (16) and Richardson and Richarson (19). The amino acid residues common to both SAP-1 and FLU NA are indicated in the one letter code for amino acids and the glycosylation site is shown as a black arrow.

X-ray crystallography revealed that FLU NA is folded in an unusual array of 6 extended β-sheets denoted β_1 to β_6 (Fig. 2A) (14). Each sheet has a "W" topology (18) with four strands connected by reverse turns (14). Considering that 22% of the FLU NA (subtype N2) amino acids are conserved in SAP-1's first 48 N-terminal amino acids, and that several amino acids are replaced conservatively, it is possible that SAP-1 adopts a secondary and tertiary structure similar to that of FLU NA in the region of homology. Fig. 2A shows the regions corresponding to SAP-1 on the 3D model of FLU NA and Fig. 2B proposes a folding structure for SAP-1 taking into account the regions of homology and a predicted C-terminal α-helix (see below).

The overall impression is that the structure of SAP-1 is more compact than the corresponding region of NA because several deletions were introduced (Fig. 1 and 2). In SAP-1, the β-sheets would correspond to β_3S_1, β_3S_2 of NA (linked by disulfide bridges), shortened β_3S_3 and β_3S_4 (with a glycosylation site in the loop between them), and β_4S_1 and β_4S_2, also linked by a disulfide bridge. Deletions are localized in the loops and the two central β-sheets and the model was constructed so that chain folding and position of functional amino acids is minimally altered with respect to FLU NA (Fig. 2B).

An α-helix is predicted at the C-terminal end (residues 56 to 65) of SAP-1 using the method of Garnier et al. (16). The helix also follows the Richardson and Richardson's rules of amino acid preferences at the N-terminal end of α-helices (19) because it contains a Ser residue at the interface followed by Glu as the first and an hydrophobic residue, Ile, as the fourth residue. In our model of SAP-1 folding (Fig. 2B), the α-helix was arbitrarily oriented parallel to β_4S_2 because, in most proteins, interacting helices and β-sheets adopt such an orientation (20). The mean hydrophobicity per residue of this helix (<H> = 0.40) is high enough to classify it as a lipid-binding helix similar to phospholipid-binding helices of apolipoproteins (<H> values between -0.005 and 0.60 as computed from phospholipid-binding peptides given in ref. 21) but is lower than strongly hydrophobic transmembrane helices of integral membrane proteins (values between 0.46 and 1.40 (17)). Also in agreement with a lipid-binding function for this helix is its amphiphilic nature. The asymmetric distribution of polar and hydrophobic amino acid residues around the helix is shown in the helical wheel projection of Fig. 3. A $<\mu_H>$ of 0.28 was computed which puts its amphiphilicity below values of surface-seeking helices. These helices are highly amphiphilic and disrupt membra-

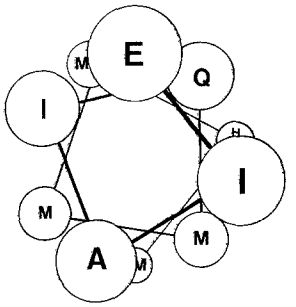

Fig. 3: Helical wheel projection of the predicted C-terminal α-helix of SAP-1. Hydrophilic amino acids are clustered on one side of the chain.

nes by interacting at the lipid-water interface (17). However, the $<\mu_H>$ value is similar to that of lipid-binding helices of apolipoproteins ($<\mu_H>$ between 0.22 and 0.43 as computed by us from data in ref. 21). The hydrophobic character and amphiphilicity of the C-terminal helix of SAP-1 appears strong enough for picking up lipids from micelles or membranes but at the same time permit rapid lipid exchanges required for the activator function.

In conclusion, this paper proposes a folding structure of SAP-1 based on its homology with FLU NA and secondary structure prediction of its C-terminal end. Although a more definite structural model of SAP-1 must await a X-ray crystallographic study of purified SAP-1, the model takes into account the existence of two structural domains in SAP-1, each corresponding to a recognition function of the activator protein: interaction with a lysosomal enzyme and sphingolipid binding. The model suggests experiments aimed at defining the mechanism of SAP-1 binding to lysosomal enzymes possibly through interaction with sialic acid residues at the surface of lysosomal enzyme and SAP-1 binding to sphingolipids by mean of an amphiphilic α-helix.

ACKNOWLEDGMENTS

The author thanks Peter M. Colman, Damian Labuda, and David A. Wenger for stimulating discussion and encouragement. The author acknowledges the gift of the protein sequence analysis software package by the Molecular Biology Computer Research Resource, Harvard University and the use of the sequence databases at l'Institut de Recherches Cliniques de Montréal, Université de Montréal. This research is supported by a grant from the Medical Research Council of Canada (MT-5163).

REFERENCES

1. Mehl. E. and Jatzkevitz, H. (1964) Hoppe-Seyler's Z. Physiol. Chem. 339, 260-276.
2. Ho, M. W. and O'Brien, J. S. (1971) Proc. Natl. Acad. Sci. USA 68, 2810-2813.
3. Li, S.-C. and Li Y.-T. (1976) J. Biol. Chem. 251, 1159-1163.

4. Hechtman, P. (1977) Can. J. Biochem. 55, 315-324.
5. Gartner, S., Conzelmann, E., Sandhoff, K. (1983) J. Biol. Chem. 258, 12378-12385.
6. Fujibayashi, S. and Wenger D. A. (1986) Biochem. Biophys. Acta 875, 554-562.
7. Inui, K., Emmett, M., Wenger, D. A. (1983) Proc. Natl. Acad. Sci. USA 80, 3074-3077.
8. Li, S.-C., Kihara, H., Serizawa, S., Li., Y.-T., Fluharty, A. L., Mayes, J. S., Shapiro, L. (1985) J. Biol. Chem.260, 1867-1871.
9. Dewji, N., Wenger D. A., Fujibayashi, S., Donoviel, M., Esch, F., Hirll F., O'Brien, J. S. (1986) Biochem. Biophys. Res. Commun. 134, 989-994.
10. Doolittle, R. F. (1986) Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences. University Science Books, Mill Valley, CA.
11. Schauer, R. (1985) Trends Biochem. Sci. 10, 357-360.
12. Shaw M. W., Lamb., R. A., Erickson, B. W., Briedis, D. J., Choppin, P. W. (1982) Proc. Natl. Acad. Sci. USA 79, 6817-6821.
13. Colman, P. M., Varghese J. N., Laver, W. G. (1983) Nature 303, 41-44.
14. Varghese, J. N., Laver, W. G., and Colman, P. M. (1983) Nature 303, 35-40.
15. Lipman, D. J. and Pearson, W. R. (1985) Science 227, 1435-1441.
16. Garnier, J., Osguthorpe, J. D., Robson, B. (1978) J. Mol. Biol. 120, 97-120.
17. Eisenberg, D., Schwartz, E., Komaromy, M. and Wall, R. (1984) J. Mol. Biol. 179, 125-142.
18. Richardson, J. S. (1977) Nature 268, 495-500.
19. Richarson, J. S. and Richardson, D. C. (1988) Science 240, 1648-1652.
20. Chothia, C. (1984) Ann. Rev. Biochem. 53, 537-572.
21. Sparrow, J. T. and Gotto, A. M. (1980) Ann. N. Y. Acad. Sci. 348, 187-208.