Homology between the human vitamin D-binding protein (group specific component), α-fetoprotein and serum albumin

Françoise Schoentgen, Marie-Hélène Metz-Boutigue, Jacqueline Jollès, Jacques Constans+ and Pierre Jollès

Laboratoire des Protéines (Unité CNRS No. 102 alliée à l'INSERM), Université de Paris V, 45 rue des Saints-Pères, F75270 Paris Cedex 06, and *+Centre d'Hémotypologie du CNRS, C.H.U. Purpan, Avenue de Grande-Bretagne, F31000 Toulouse, France*

Received 1 April 1985

41 Amino acid long N-terminal sequences of the three major human vitamin D-binding proteins (group-specific components Gc1F, Gc1S and Gc2) were characterized: they were identical. By computer analyses, the alignment of this N-terminal sequence with several sequences of human serum pre-proalbumin and human pre-α-fetoprotein was established

Vitamin D-binding protein *Group-specific component* *Serum albumin* *α-Fetoprotein* *Amino acid sequence Homology*

1. INTRODUCTION

Discovered by Hirschfeld in 1959 [1], group-specific component (Gc) is a protein present in human plasma at a concentration of 20–50 mg/100 ml [2]. Recent interest comes from the observation that its biological role is the transport of vitamin D metabolites in plasma [3]. In addition, tissue actin binds Gc protein with high affinity [4,5] and the latter is involved in the linkage between membrane immunoglobulin (MIg) of B lymphocytes and actin [6].

Gc protein is highly polymorphic and at the present time a total of 84 different Gc alleles have been identified [7]. Nevertheless, only 3 major Gc alleles (Gc^{1F}, Gc^{1S} and Gc^{2}) have been detected in most of the populations studied so far.

Previously, the first 20 amino acid residues of the 3 major group-specific components Gc1F, Gc1S and Gc2 have been reported: their N-terminal sequences were identical [8]. In this note we report the extension of these N-terminal sequences (residues 1→41) and by computer analyses we demonstrate for the first time a structural homology between Gc proteins, human serum albumin and human α-fetoprotein.

2. MATERIALS AND METHODS

2.1. *Materials*

Human Gc proteins (Gc1F, Gc1S, Gc2) were purified from plasma fractions of individual homozygotes according to the procedure of Viau et al. [9]. All reagents (analytical grade) were purchased from Prolabo or Merck except those employed for the sequencer which were obtained from Pierce or SDS (Marseilles).

2.2. *Methods*

The 3 major Gc proteins were reduced with 2-mercaptoethanol [10] and alkylated with iodoacetamide [11]. Automated Edman degradation was carried out in a Beckman Sequencer 890 C by the 1 M quadrol method in the presence of polybrene [12]; the phenylthiohydantoin amino acids were identified by high-performance liquid

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

chromatography with a Waters chromatograph (model ALC/GPC 204) [13].

The search of sequences homologous to the N-terminal sequence of Gc proteins was performed according to the program SEARCH. The sequence data base of the Atlas of Protein Sequence and Structure updated to August 1984 as well as the program SEARCH were provided by the National Biomedical Research Foundation. Amino acid sequences were compared using the proportional matching option of the interactive graphics program DIAGON [14]. The latter incorporates a scoring system based on the MDM_{78} matrix, which was calculated from accepted point mutations of 71 families of related proteins [15]. Thus, sequences which are related in the proteins under comparison appear as diagonal lines on the graphical output. To obtain the alignment of two homologous sequences the program of Smith and Waterman [16] was applied.

3. RESULTS AND DISCUSSION

The 41 amino acid long N-terminal sequences of reduced and alkylated Gc1F, Gc1S and Gc2 proteins were automatically established (fig.1): they were identical. Two differences (Glu/Gln and Asp/Asn, in positions 2 and 6, respectively) were, however, noted between our sequence and the shorter (1→20 residues) previously proposed [8]. Based on our new sequence data we performed a first series of computer analyses (program SEARCH): they pointed out similarity between the N-terminal sequence of Gc proteins and human, bovine and rat pre-proalbumins (sequence data in [17–20]).

Previously, structural homology was clearly demonstrated between α-fetoprotein and serum albumin of different species [21–24]. To show a possible alignment between the N-terminal sequence of Gc proteins and human α-fetoprotein, we performed a second series of computer tests using the program DIAGON [14] and the method of Smith and Waterman [16]. We obtained the alignment of our N-terminal sequence with several sequences of human pre-proalbumin (pre-proHSA) [17] as well as with human pre-α-fetoprotein (pre-AFP) [21]. We report in fig.1 the structural relationship observed between the N-terminal sequence of Gc proteins and two sequences of pre-proHSA (residues 20→60 and 404→444) and one sequence of pre-AFP (residues 404→444). For these 3 alignments the sequence homology (identical and homologous residues in corresponding positions) is 34%.

In addition a clear demonstration of these sequence homologies is shown in fig.2a–c by the diagonal plot method of Staden [14]. Taking into account the strong relatedness between serum albumin and α-fetoprotein and the localization of their disulfide bridges, these two proteins could be subdivided into 3 similar repeat units called domains (I–III) [19,21–24]. The N-terminal Gc sequence corresponds to the propeptide of pre-proHSA lengthened by the beginning of domain I of HSA. Similarly, our sequence is aligned with the homologous sequences of pre-proHSA and pre-AFP (residues 404→444) which surround the beginning of domain III of each protein [21]. In a previous study, a computer search including only the 20 first amino-terminal residues of Gc proteins (program SEARCH) failed to show similarity with any reported sequence [8]. Indeed, the maximum homology is situated only after residue number 20.

Recently an immunological cross-reactivity was demonstrated between Gc protein and human

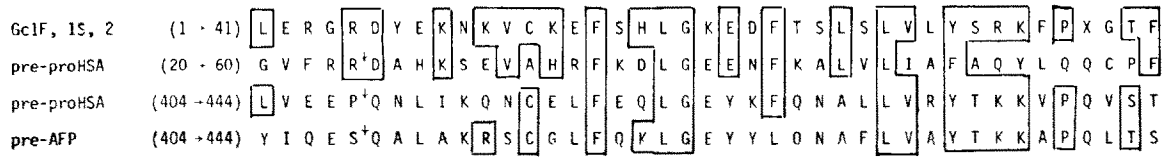

Fig.1. The N-terminal sequence of Gc proteins (residues 1→41): comparison with human pre-proalbumin (pre-proHSA) [17] and human pre-α-fetoprotein (pre-AFP) [21]. The alignment was determined according to the program of Smith and Waterman [16]. Boxed areas correspond to residues identical or homologous to the Gc protein sequence (□). The one-letter amino acid abbreviation system is used. Arrows (↓) indicate the beginning of domains I and III of human serum albumin and α-fetoprotein according to [21].

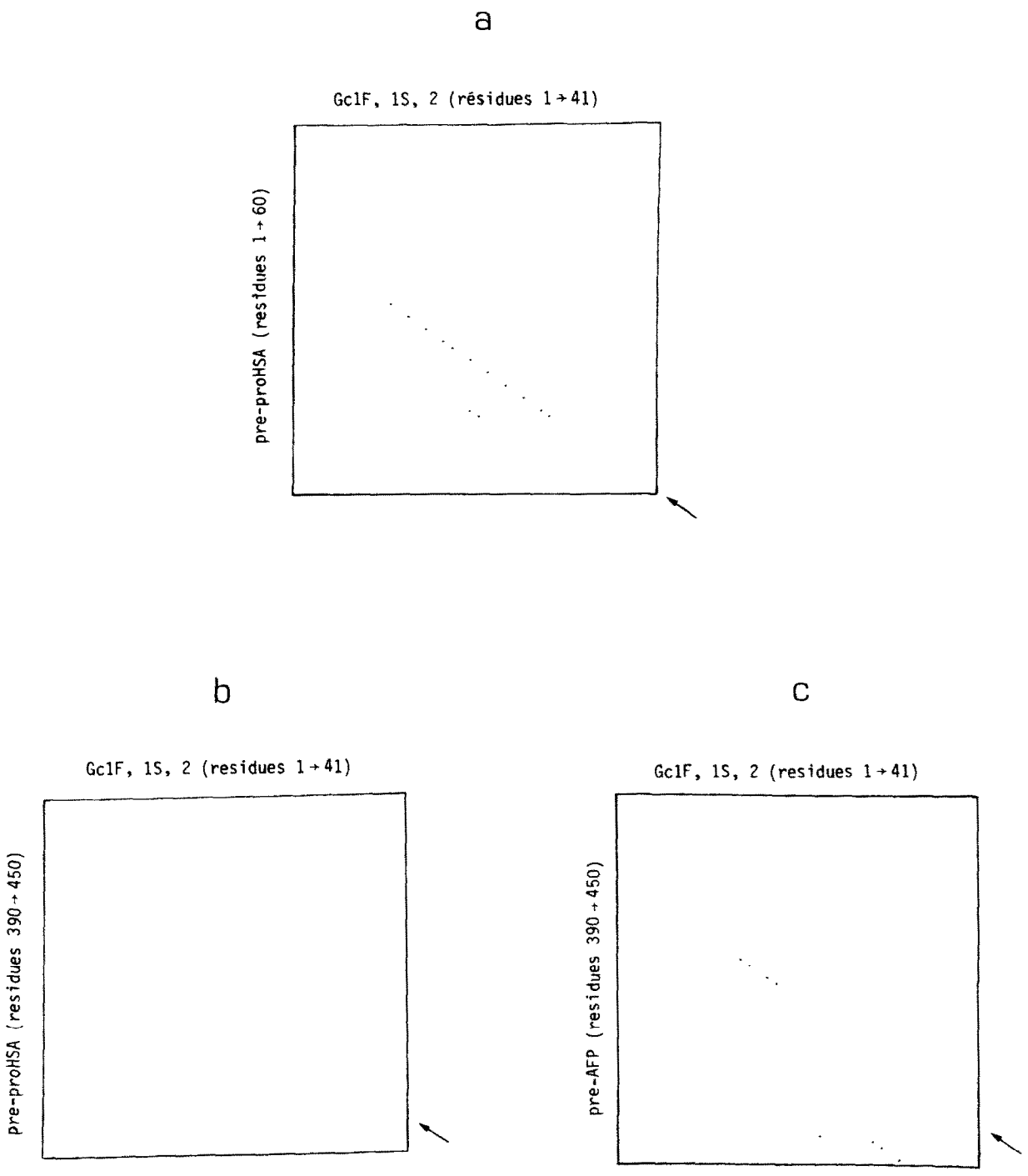

Fig.2. Diagonal plot of the N-terminal sequence of Gc proteins (residues 1→41) against human pre-proalbumin (pre-proHSA) [17] and human pre-α-fetoprotein (pre-AFP) [21] according to the program DIAGON [14]. The homology is shown by the strong diagonal clustering of points (arrow ←). (a) Diagonal plot against pre-proHSA (residues 1→60). (b) Diagonal plot against pre-proHSA (residues 390→450). (c) Diagonal plot against pre-AFP (residues 390→450).

serum albumin and human α-fetoprotein, suggesting that these 3 proteins have evolved from the same ancestral protein [25]. Our structural molecular studies reported here seem to corroborate this hypothesis.

ACKNOWLEDGEMENTS

The excellent technical assistance of Mr Ly Quan Le is gratefully acknowledged. We also thank Mr C. Mugnier and Mr C. Fondrat (CITI 2, 45 rue des Saints-Pères, 75270 Paris Cedex 06) for computer analyses.

REFERENCES

[1] Hirschfeld, J. (1959) Acta Pathol. Microbiol. Scand. 47, 160–168.
[2] Putnam, F.W. (1977) in: The Plasma Proteins (Putnam, F.W. ed.) vol.3, pp.333–357, Academic Press, New York.
[3] Daiger, S.P., Schanfield, M.S. and Cavalli-Sforza, L.L. (1975) Proc. Natl. Acad. Sci. USA 72, 2076–2080.
[4] Haddad, J.G. (1982) Arch. Biochem. Biophys. 213, 538–544.
[5] Coué, M., Constans, J., Viau, M. and Olomucki, A. (1983) Biochim. Biophys. Acta 759, 137–145.
[6] Petrini, M., Emerson, D.L. and Galbraith, R.M. (1983) Nature 306, 73–74.
[7] Constans, J., Cleve, H., Dykes, D., Fischer, M., Kirk, R.L., Papiha, S.S., Scheffran, W., Scherz, R., Thymann, M. and Weber, W. (1983) Hum. Genet. 65, 176–180.
[8] Svasti, J., Kurosky, A., Bennett, A. and Bowman, B.H. (1979) Biochemistry 18, 1611–1617.
[9] Viau, M., Constans, J., Debray, H. and Montreuil, J. (1983) Biochem. Biophys. Res. Commun. 117, 324–331.
[10] Crestfield, A.M., Moore, S. and Stein, W.H. (1963) J. Biol. Chem. 238, 622–627.
[11] Jollès, J., Van Lemputten, E., Mouton, A. and Jollès, P. (1972) Biochim. Biophys. Acta 257, 497–510.
[12] Tarr, G.E., Beecher, J.F., Bell, M. and McKean, D.J. (1978) Anal. Biochem. 84, 622–627.
[13] Pellicone, C., Bouillon, P., Virmaux, N. and Vincendon, G. (1981) Biochimie 63, 671–676.
[14] Staden, R.A. (1982) Nucleic Acids Res. 10, 2951–2961.
[15] Schwartz, R.M. and Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure, vol.5, suppl.3, pp.353–358, National Biomedical Research Foundation, Washington.
[16] Smith, T.F. and Waterman, M.S. (1981) Adv. Appl. Math. 2, 482–489.
[17] Dugaiczyk, A., Law, S.W. and Dennison, O.E. (1982) Proc. Natl. Acad. Sci. USA 79, 71–75.
[18] MacGillivray, R.T.A., Chung, D.W. and Davie, E.W. (1979) Eur. J. Biochem. 98, 477–485.
[19] Brown, J.R. (1976) Fed. Proc. 35, 2141–2144.
[20] Sargent, T.D., Yang, M. and Bonner, J. (1981) Proc. Natl. Acad. Sci. USA 78, 243–246.
[21] Morinaga, T., Sakai, M., Wegmann, T.G. and Tamaoki, T. (1983) Proc. Natl. Acad. Sci. USA 80, 4604–4608.
[22] Jagodzinski, L.L., Sargent, T.D., Yang, M., Glackin, C. and Bonner, J. (1981) Proc. Natl. Acad. Sci. USA 78, 3521–3525.
[23] Gorin, M.B., Cooper, D.L., Eiferman, F., Van de Rijn, P. and Tilghman, S.M. (1981) J. Biol. Chem. 256, 1954–1959.
[24] Law, S.W. and Dugaiczyk, A. (1981) Nature 291, 201–205.
[25] Pekkala-Flagan, A. and Ruoslahti, E. (1982) J. Immunol. 128, 1163–1167.