Vol. 149, No. 1, 1987 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1987 Pages 282-288

THE HOMOLOGY OF COMPLEMENT FACTOR C8 GAMMA CHAIN AND ALPHA-1-MICROGLOBULIN

Lois T. Hunt, Andrzej Elzanowski, and Winona C. Barker

Protein Identification Resource
National Biomedical Research Foundation
Georgetown University Medical Center
3900 Reservoir Road, N.W.
Washington, D.C. 20007

Received October 13, 1987

SUMMARY. The sequence of the complement factor C8 gamma chain shares a remarkable degree of similarity with that of alpha-1-microglobulin, a member of the alpha-2u-globulin superfamily. This superfamily comprises a diverse group of distantly related animal proteins possessing characteristic structural features and similar functions. Comparison of the C8 gamma chain with these proteins supports its homology to them and suggests a possible functional role.

The human complement factor C8 molecule is one component of the plasma cytolytic C5b-9 complex, also termed the membrane attack complex (reviewed by Muller-Eberhard [1]). In the C8 molecule the disulfide-linked alpha-gamma dimer is noncovalently associated with the beta chain. The cDNA sequences of all three chains have now been determined (2-4), and it has been conclusively shown that each is the product of a single gene (2,4). Although domains and features of the alpha and beta proteins have been described and correlated with known or proposed functions ([4] and references therein from the same laboratory), the role of the gamma chain is still obscure (4,5). This protein was described as having no obvious sequence features and no homologies to any other protein (4).

However, our visual inspection of the C8 gamma chain revealed to us an array of features that indicated a probable relationship to proteins in the alpha-2u-globulin superfamily, which we are reviewing in more detail elsewhere (6). These features included the length of the chain, the number and position of cysteines, and the presence of three 3-residue motifs (N-F-D, G-T-W, T-D-Y) at the appropriate positions in the chain, that are similar to or identical with those in proteins of this superfamily. Computer analyses confirmed the relationship and allowed us to make certain predictions about structural features and suggestions about functional roles of the C8 gamma chain.

0006-291X/87 $1.50

RESULTS OF COMPUTER ANALYSES AND DISCUSSION

A search of the Protein Sequence Database of the Protein Identification Resource (7), using the FASTP program of Lipman and Pearson (8), revealed a top-scoring match with human alpha-1-microglobulin; the initial and optimized scores were 95 and 209, respectively, indicating a probable relationship (8). Alpha-1-microglobulin (also known as protein HC) and inter-alpha-trypsin inhibitor are amino and carboxyl domains cleaved from a common precursor (9). Several other proteins in this same superfamily, such as plasma retinol-binding protein (10-12), epididymal androgen-dependent protein (13), and beta-lactoglobulin (14,15), also obtained optimized scores high enough to indicate a possible relationship. Some of the more distantly related proteins in this superfamily (6) were not found in the search.

Several proteins have recently been added to the alpha-2u-globulin superfamily (16,17) and it now includes at least 11 different proteins, based in most cases primarily on sequence evidence (reviewed in [6]). The mature proteins are of approximately the same lengths, ranging from 162 to 189 residues. The homology of rat alpha-2u-globulin and bovine beta-lactoglobulin was the first to be reported (18), although the relationships of human plasma retinol-binding protein and human alpha-1-microglobulin were also detected by computer analyses we performed at the time for P. Feigelson. Among more recent additions to the superfamily were human apolipoprotein D and tobacco hornworm insecticyanin (19,20), human epididymal androgen-dependent protein (21), chicken purpurin (22), frog olfactory protein (23), and human and rodent alpha-1-acid glycoprotein (6,16). All of these proteins are secreted into body fluids and most are known to bind small lipophilic molecules; some also bind carbohydrate or associate with other molecules (6,16).

The sequence of C8 gamma chain was compared in two ways with those of the nine complete protein sequences already included in this superfamily. Dotmatrix graphic plots were generated to reveal the regions of similarity, using the Mutation Data Matrix (24) with a minimum score of 20 and a window size (segment length) of 20. Statistical estimates of sequence similarities were obtained with our ALIGN program (24), using the Mutation Data Matrix with a bias of 6, a gap penalty of 6, and 100 randomizations of the sequence pairs.

The greatest degree of similarity was found between the C8 gamma chain and alpha-1-microglobulin; the dotmatrix plot shows a diagonal line, with only a few short gaps, over the entire lengths of both sequences (see Figure 1) and the ALIGN score is over 13 SD (see Figure 2). If the signal sequences are

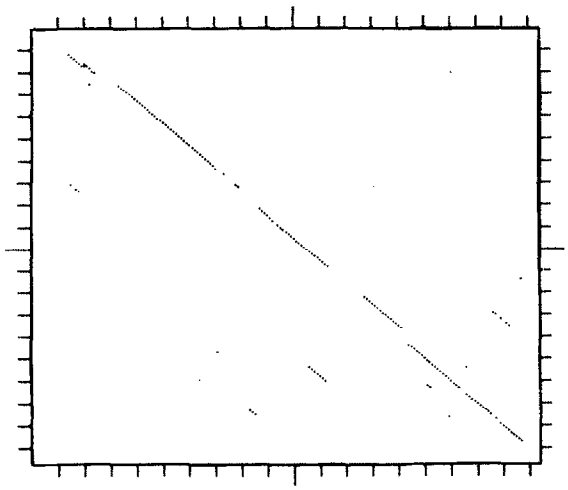

Figure 1. Dotmatrix graphic plot of human complement factor C8 gamma chain precursor (horizontal axis) and human alpha-1-microglobulin precursor (vertical axis). The diagonal lines represent areas of similarity. A segment length of 20 residues and a minimum score of 20 with the Mutation Data Matrix were used in the comparison.

excluded, the two proteins are only 25-27% identical (depending on the positioning of gaps), but when conservative substitutions are counted there are similar residues at over 40% of the remaining positions (about 70% total similarity). Conserved residues in the proteins of this superfamily tend to be clustered in several short regions. This can be seen in the alignment of the two proteins in Figure 2: for example, positions 41-63, which include a conserved cysteine and the N-F-D/N-F-N and G-T-W/G-R-W motifs (the Trp is conserved in all the proteins in the superfamily); positions 86-98 and 181-195, which contain the two cysteines that form the conserved disulfide bond (Cys-195 is conserved in all of the proteins), as well as two sites in at least three of the proteins corresponding to two intron positions in the gene sequences; and positions 153-172, which also include an intron position in at least two of the proteins (see [6] and references therein). Based on the comparison of the three cysteines in the C8 gamma chain with cysteines and established disulfide bonds in the proteins of this superfamily (6), we predict that the second and third cysteines will form an intrachain bond, leaving the first cysteine to form the interchain bond with the C8 alpha chain.

Dotmatrix plots and ALIGN scores for comparisons of the C8 gamma chain with other proteins in the superfamily indicate less similarity. To eliminate length differences at the amino and carboxyl ends among the proteins, the sequences used in ALIGN comparisons corresponded to positions 37-197 in the alignment of Figure 2. The score was 13.2 Sd for the C8 gamma chain and alpha-1-microglobulin; the scores with olfactory protein, plasma retinol-

```
          1        10        20        30        40        50        60        70
          '        '         '         '         '         '         '         '
CF8G      MLPPGTATLLTLLLAAGSLGQKPQRPRRPASPISTIQPKANFDAHQFAGTWLLVAVGSACRFLQEQGHRAE
          :   :   :: ::: :  :          :  :   ::   ::       : :   : :: :  :      :
A1MG      MRSLGA--LL-LLLSA-CLAVSAGPV--PTPP-DNIQVQENFNISRIYGKWYNLAIGSTCPWLKKIMDRMT

         72        80        90       100       110       120       130       140
          '        '         '         '         '         '         '         '
CF8G      ATTLHVAPQGTA--MAVSTFRKLDGICWQVRQLYGDTGVLGRFLLQARGARGAVHVVVAETDYQSFAVL--
            ::      :         :   : :     :   :   : ::             : : :   :
A1MG      VSTLVLGEGATEAEISMTSTRWRKGVCEETSGAYEKTDTDGKFLYHKSKWNITMESYVVHTNYDEYAIFLT
                                    *
        143     150       160       170       180       190       200       210
          '        '         '         '         '         '         '         '
CF8G      --YLERAGQ-LSVKLYARSLPVSDSVLSGFEQRVQEAHLTEDQIFYFPKYGFCEAADQFH--VL-DEVRR
             :    ::: :      :  :    :    :: ::    : :   :     :    :::
A1MG      KKFSRHHGPTITAKLYGRAPQLRETLLQDFRVVAQGVGIPEDSIFTMADRGECVPGEQEPEPILIPRVRR
                                                    *
```

```
Total score = 1445, 10 breaks
  57 identities out of 195 possible matches between residues

100 random runs
Alignment score =   13.04 SD   Standard deviation =   19.30   Mean = 1193.30
```

Figure 2. ALIGN of human complement factor C8 gamma chain precursor and human alpha-1-microglobulin precursor. Arrows indicate signal sequence cleavage sites. Cysteines forming an intrachain disulfide bond in alpha-1-microglobulin are starred.

binding protein, epididymal androgen-dependent protein, beta-lactoglobulin, and alpha-2u-globulin were 9.6, 8.6, 8.1, 8.0, and 7.1 SD, respectively. The scores with human and rat alpha-1-acid glycoprotein were 5.1 and 4.8 SD, and with apolipoprotein D and insecticyanin, 4.1 and 5.1 SD. The ALIGN score for the C8 gamma chain and alpha-1-microglobulin was one of the highest for any pair of proteins in the superfamily. As the other proteins are held to have a common evolutionary origin, there can be no doubt that the C8 gamma chain is homologous to all of them.

The HPLOT, ALOM, and CHOFAS programs of Kanehisa (25) were used to generate a hydrophobicity plot, to detect potential membrane-spanning segments, and to predict secondary structure for the C8 gamma chain. In agreement with Ng et al. (4), no potential transmembrane regions were found. Two segments, including residues 24-40 and 52-69 in the mature protein, are moderately hydrophobic but each contains a Glu and a Ser and they are barely long enough to cross a membrane. However, none of the other proteins in this superfamily has a membrane-spanning domain. Comparison of the CHOFAS prediction (even though this may be only 50-60% accurate) with the known three-dimensional structures of beta-lactoglobulin (26) and plasma retinol-binding protein (27) shows that the C8 gamma chain could possess a similar structure, in which eight successive antiparallel beta sheets, folded into a beta barrel, are followed by a short alpha-helix. The areas in the C8 gamma chain that correspond to the positions of the beta sheets and helix in beta-lactoglobulin are

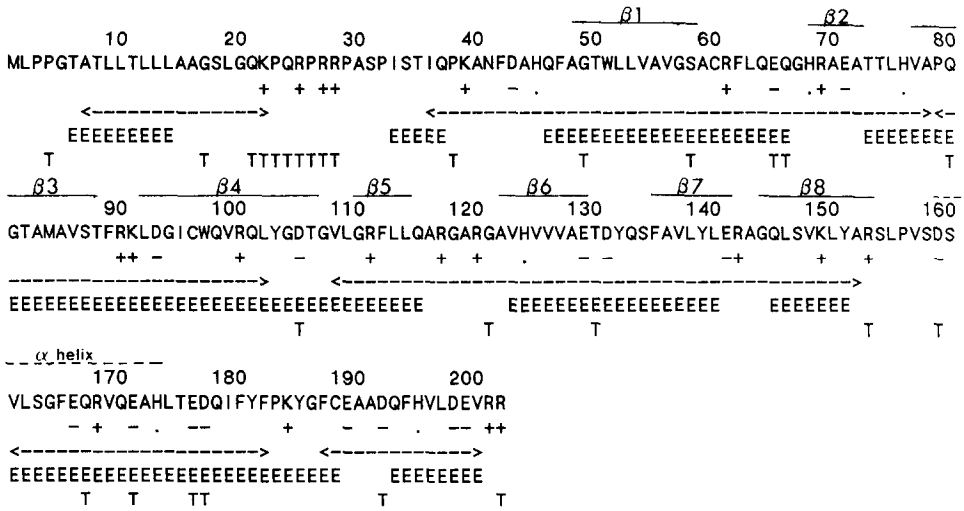

Figure 3. CHOFAS prediction of secondary structure for human complement factor C8 gamma chain. Segments corresponding to the eight beta sheets and the alpha helix of beta-lactoglobulin are indicated above the sequence. Below the sequence, hydrophilic charged residues are marked by + or - and hydrophobic histidines, by dots; alpha helix, by <-->; beta sheet, by EE; beta turn, by T.

indicated in Figure 3; only the short second beta sheet is not predicted. Turns are also predicted approximately at the ends of many of the beta sheets.

The beta barrel structure, which forms a cup-shaped hydrophobic pocket, is correlated in most of these proteins with the binding of small lipophilic molecules, such as retinol (27), the yellow-brown chromophore reported for alpha-1-microglobulin (28), or some type of lipid, as reported for human apolipoprotein D (19,20) and alpha-1-acid glycoprotein (29). The C8 gamma chain, which appears to have the capability to form the same structure, could also possess such a function, although we could find no report of an association of the gamma chain with any small lipophilic molecule. The gamma chain is not essential for the activity of the C8 molecule in the C5b-C9 complex yet it enhances binding of the C8 beta chain to the C5b-C7 complex, either directly by interaction with C5b-7 or indirectly through its associations with the C8 alpha chain (5). In addition to the interchain disulfide bond, there is a noncovalent association of the gamma chain and the alpha chain. This interaction of proteins within a complex is reminiscent of the association of the related retinal protein purpurin with the neural cell adhesion molecule in retinal adherons (22); purpurin, which can also bind retinol, stimulates adhesion by interacting with heparan sulfate proteoglycan on the surface of the retinal cells. Some possible roles for the C8 gamma chain have been suggested (4), including participation in protecting homologous cells from lysis by the complement membrane attack complex (1,4) through interaction with a protein on the cell surface.

However, we can envision at least two other possible roles for the C8 gamma chain, in the light of this newly discovered relationship. One is for the gamma chain to bind and transport a small lipophilic molecule, either when the C8 molecule is free in the plasma or when it is in the C5b-C9 complex. The second, more likely, possibility is that its association with the C8 alpha chain and its potential for binding are correlated; that is, the gamma chain associates noncovalently with some type of exposed structural feature on the alpha chain, such as a small hydrophobic finger or disulfide loop.

ACKNOWLEDGMENTS

This research was supported by grants RR01821, from the Division of Research Resources, and CA40474, from the National Cancer Institute, of the National Institutes of Health.

REFERENCES

1. Muller-Eberhard, H.J., Annu. Rev. Immunol. 4: 503-528, 1986
2. Rao, A.G., Howard, O.M.Z., Ng, S.C., Whitehead, A., Colten, H.R., and Sodetz, J.M., Biochemistry 26: 3556-3564, 1987
3. Howard, O.M.Z., Rao, A.G., and Sodetz, J.M., Biochemistry 26: 3565-3570, 1987
4. Ng, S.C., Rao, A.G., Howard, O.M.Z., and Sodetz, J.M., Biochemistry 26: 5229-5233, 1987
5. Brickner, A., and Sodetz, J.M., Biochemistry 24: 4603-4607, 1985
6. Elzanowski, A., Hunt, L.T., and Barker, W.C., manuscript in preparation
7. Protein Sequence Database of the Protein Identification Resource (PIR), Release 13.0, June 30, 1987, National Biomedical Research Foundation, Washington, DC
8. Lipman, D.J., and Pearson, W.R., Science 227: 1435-1441, 1985
9. Kaumeyer, J.F., Polazzi, J.O., and Kotick, M.P., Nucl. Acids Res. 14: 7839-7850, 1986
10. Colantuoni, V., Romano, V., Santoro, C., Bensi, G., Costanzo, F., Raugei, G., and Cortese, R., Nucl. Acids Res. 11: 7769-7775, 1983
11. Sundelin, J., Anundi, H., Tragardh, L., Eriksson, U., Lind, P., Ronne, H., Peterson, P.A., and Rask, L., J. Biol. Chem. 260: 6488-6493, 1985
12. Laurent, B.C., Nilsson, M.H.L., Bavik, C.O., Jones, T.A., Sundelin, J., and Peterson, P.A., J. Biol. Chem. 260: 11476-11480, 1985
13. Brooks, D.E., Means, A.R., Wright, E.J., Singh, S.P., and Tiver, K.K., J. Biol. Chem. 261: 4956-4961, 1986
14. Braunitzer, G., Chen, R., Schrank, B., and Stangl, A., Hoppe-Seyler's Z. Physiol. Chem. 354: 867-878, 1973
15. Preaux, G., Braunitzer, G., and Kolde, H.-J., Arch. Int. Physiol. Biochim. 88: B45-B46, 1980
16. Pervaiz, S., and Brew, K., FASEB J. 1: 209-214, 1987
17. Sawyer, L., Nature 327: 659, 1987
18. Unterman, R.D., Lynch, K.R., Nakhasi, H.L., Dolan, K.P., Hamilton, J.W., Cohn, D.V., and Feigelson, P., Proc. Nat. Acad. Sci. USA 78: 3478-3482, 1981
19. Drayna, D., Fielding, C., McLean, J., Baer, B., Castro, G., Chen, E., Comstock, L., Henzel, W., Kohr, W., Rhee, L., Wion, K., and Lawn, R., J. Biol. Chem. 261: 16535-16539, 1986
20. Drayna, D.T., McLean, J.W., Wion, K.L., Trent, J.M., Drabkin, H.A., and Lawn, R.M., DNA 6: 199-204, 1987
21. Brooks, D.E., Biochem. Int. 14: 235-240, 1987

22. Schubert, D., LaCorbiere, M., and Esch, F., J. Cell Biol. 102: 2295-2301, 1986
23. Lee, K.-H., Wells, R.G., and Reed, R.R., Science 235: 1053-1056, 1987
24. Dayhoff, M.O., Barker, W.C., and Hunt, L.T., Methods Enzymol. 91: 524-545, 1983
25. Kanehisa, M., IDEAS (Integrated Database and Extended Analysis System for Nucleic Acids and Proteins) User Manual, Nat. Cancer Inst., Nat. Inst. Health, 1985
26. Papiz, M.Z., Sawyer, L., Eliopoulos, E.E., North, A.T.C., Findley, J.B.C., Sivaprasadarao, R., Jones, T.A., Newcomer, M.E., and Kraulis, P.J., Nature 324: 383-385, 1986
27. Newcomer, M.E., Jones, T.A., Aqvist, J., Sundelin, J., Eriksson, U., Rask, L., and Peterson, P.A., EMBO J. 3: 1451-1454, 1984
28. Gavilanes, J.G., Lopez-Otin, C., Gavilanes, F., and Mendez, E., Biochemistry 23: 1234-1238, 1984
29. Kute, T., and Westphal, U., Biochim. Biophys. Acta 420: 195-213, 1976