# REPETITIVE SEQUENCES IN PROTEIN A FROM *STAPHYLOCCUS AUREUS*: THREE HIGHLY HOMOLOGOUS Fc-BINDING REGIONS

Jörgen SJÖDAHL
*Department of Medical and Physiological Chemistry, The Biomedical Center, University of Uppsala, Box 575, S-751 23 Uppsala, Sweden*

Received 29 April 1976

## 1. Introduction

Protein A, part of the cell wall structure of *Staphylococcus aureus*, is covalently linked to the peptidoglycan moiety of the cell wall [1]. Its mol. wt. is 42 000, the frictional ratio 2.1–2.2 and the intrinsic viscosity 29 ml/g [2]; this suggests a markedly extended shape.

Protein A reacts with the Fc-region of IgG from several species [3–5] and is able to bind two or more molecules of IgG per molecule [6]. This interaction has been utilized in methods for immunochemical and cell surface structure studies [7,8]. Partial tryptic digestion of the protein and subsequent isolation of Fc-binding products by affinity chromatography and by ion exchange chromatography on IgG-Sepharose and phosphocellulose columns respectively showed that the native molecule exhibits at least three Fc-binding regions. Furthermore, it was suggested that these regions were similar with respect to the amino acid sequences. Digestion of intact bacteria could also be used as a preparative method for fragments similar to those obtained by digestion of isolated protein [9].

In the present work the Fc-binding products have been further purified and from the N-terminal amino acid sequence data it is possible to assign the fragments to three highly homologous Fc-binding regions, each consisting of more than 50 amino acid residues. In the sequenced parts, 80% of the residues are identical when the regions are mutually compared.

## 2. Materials and methods

### 2.1. *Preparation and isolation of Fc-binding protein A fragments*

Tryptic Fc-binding fragments with mol. wts. around 7000 were prepared as described in [9]. The isolation by chromatography on IgG-Sepharose and phosphocellulose columns were performed essentially as described in [9]. After desalting on Sephadex G-25 (Pharmacia Fine Chem. AB, Sweden) in 0.05 M $NH_4OH$ the isolated products were applied to DEAE-cellulose columns (Whatman DE 32, Whatman Biochemicals Ltd, England) in 0.05 M $NH_4OH$ and eluted with a linear $NH_4Cl$ gradient at pH 9.2. Final desalting was performed on Sephadex G-25 in 1 mM HCl.

### 2.2. *Analytical methods*

Immunochemical studies were performed essentially as described in [9].

Amino acids were analysed according to Spackman et al. [10] on a Beckman 121-M amino acid analyzer and quantified by a Hewlett-Packard electronic integrator (Mod. 3371 B). Generally, 3 nmol of samples were hydrolysed in 6 M HCl (1% phenol, w/v) at 110°C for 24 and 72 h. Threonine and serine contents were extrapolated to zero time. Valine and isoleucine were determined from the 72 h hydrolyses.

N-terminal amino acid determinations were performed by the dansyl technique [11]. Dansyl chloride was purchased from BDH Chemicals Ltd, England and the polyamide thin-layer sheets used for


*North-Holland Publishing Company – Amsterdam*

identification [12] from Cheng Chin Trading Co., Taiwan.

Amino acid sequences were analysed manually [13] or with the Beckman Sequencer mod. 890 using a modified DX-10 program. Approximately 300 nmol samples were used for sequencing. In some cases, the sample was treated with sulphonated naphthylisothiocyanate (Braunitzer reagent IV, [4]) before the sequencer run. The PTH-amino acids were separated on polyacrylamide thin-layer sheets [15] and detected by u.v. absorption [16]. Isoleucine and leucine were distinguished by amino acid analysis after hydrolysis in 6 N HCl, 0.1% $SnCl_2$ at 150°C for 4 h [17] and histidine and arginine by specific staining with the Pauly [18] and a modified Sakaguchi [19] reagent.

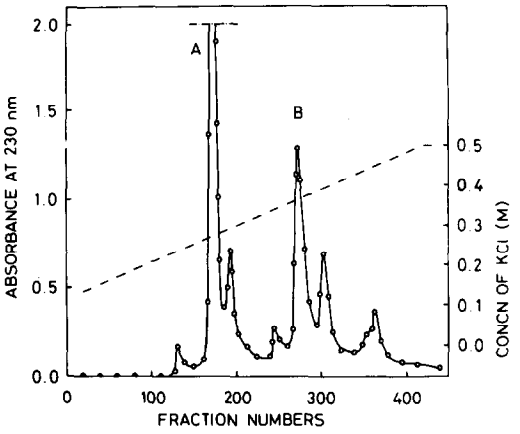


Fig.2. Phosphocellulose chromatogram from the separation of active material obtained by partial tryptic digestion of *S. aureus*. Experimental details as described in fig.1.

## 3. Results and discussion

Figs.1 and 2 illustrate the elution profile from the phosphocellulose column separations of Fc-binding material generated by partial tryptic digestions of isolated protein A and of intact bacteria respectively [9]. Chromatography on DEAE-cellulose columns showed that fragment A, II, VI and VII could be further separated while fragment B, I. III, IV and V gave single peaks. The DEAE-chromatograms of

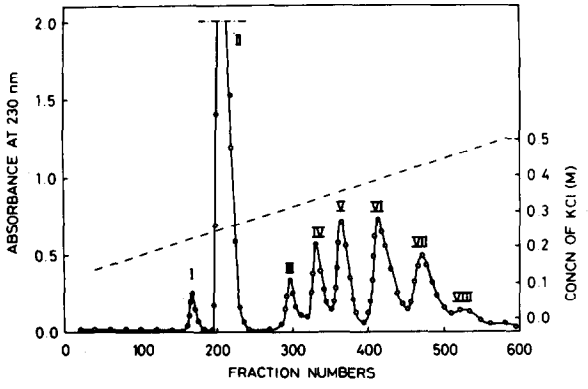


Fig.1. Separation of active tryptic protein A fragments by ion-exchange chromatography on a phosphocellulose column. Flow rate 125 ml/h; fraction volume 5 ml. Gradient 0.1–0.5 M KCl in 0.025 N $H_3PO_4$, 0.13 mM/ml. Column size 2 × 40 cm.

fragment II, VI and VII are shown in fig.3. The fragment A chromatogram was identical to that of fragment II, and fragment B gave the same chromatogram as fragment IV. Because of this and the data presented in [9], fragments A and II and fragments B and IV respectively were considered to be identical products, hence fragments A and B are omitted in the following characterizations.

Fragment VI was split into three compounds VI′, VI″ and VI‴. However, no significant difference between them could be shown in the characterizations described below. The same is true also for fragment VII. In the following only the fragments obtained in the highest yields namely VI″ and VII‴ are considered.

The N-terminal and the amino acid analyses of the isolated material are presented in table 1. The formula weights, calculated from the estimated number of residues are included. Mol. wt. determinations by gel chromatography and gel electrophoresis performed on the phosphocellulose material ranged from 5500–7500 [9]. The N-terminal analyses, the homogeneity in polyacrylamide gel electrophoresis on 11% gels at pH 8.9 [9], the homogeneity in both cation and anion exchange chromatography, and the amino acid analysis data all indicate that pure fragments were obtained.

The results from the immunochemical tests are in complete agreement with those in [9], i.e. all fragments gave reactions of mutual identity when tested

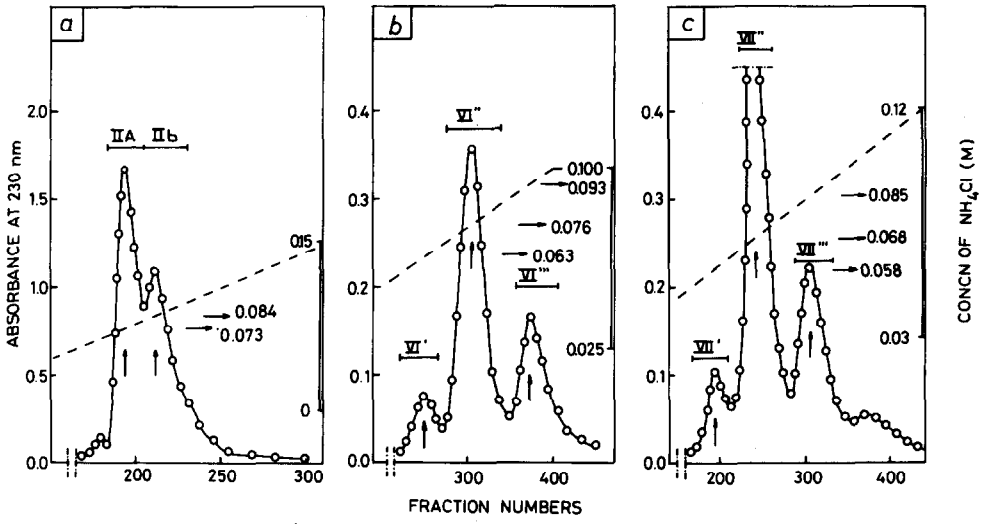


Fig.3. Further separation of active tryptic protein A fragments by ion exchange chromatography on DEAE-cellulose columns. (a) Fragment II; column 1.5 × 9 cm; Flow rate 62 ml/h; Fraction volume 1 ml; gradient 0–0.15 M $NH_4Cl$ pH 10–9.2, 0.50 mM/ml. (b) Fragment VI; column 1.5 × 10 cm; flow rate 60 ml/h; Fraction volume 1 ml; gradient 0.025–0.100 M $NH_4Cl$ pH 9.2, 0.19 mM/ml. (c) Fragment VII; Column 1.5 × 10 cm; flow rate 71 ml/h; fraction volume 1 ml; Gradient 0.03–0.12 M $NH_4Cl$ pH 9.2, 0.20 mM/ml.

against rabbit anti-protein-A serum and against pooled human IgG. Partial identity with protein A was observed against the anti-protein A serum. The hemagglutination and hemagglutination inhibition tests of the fragments point towards the presence of *one* binding site per molecule. The Fc-binding activities relative to that of protein A are given in table 1.

From the data presented above the working hypothesis would be that the fragments originate from three slightly different Fc-binding regions in protein A. Fragment I and IIA, fragment IIb, III and IV, and fragment V, VI'' and VII'' should be generated from three regions in protein A, designated region A, B, and C. Accordingly, the fragments are classed into three groups A, B, and C in table 1.

To verify the hypothesis, it is necessary to sequence all the fragments throughout the entire chain. However, N-terminal sequences of fragments from the various groups would constitute very strong support for the hypothesis provided that the sequences really show differences. Fig.4 illustrates the data from the sequence analyses of two fragments from each group, namely fragment I, IIA, IIb, IV, V, and VII''.

The sequencing method used for the various samples are noted in the figure. If the sequences and the amino acid compositions are mutually compared the classing of the fragments into three groups seems adequate. From the experimental data the amino acid sequences of the corresponding regions in protein A have been suggested. Region A has an Asn in position 4 and 18, and a Met in position 19. Region B differs in this respect by exhibiting a Lys in position 4, a His in position 18, and a Leu in position 19. Considering region C, the only established difference compared to region B shows up in position 23 (according to the alignment). The amino acid occupying this position was shown to be a Thr in region C and an Asn in region B. If one calculates the amino acid compositions of the residual peptide parts by substracting the sequenced parts from the determined total amino acid compositions, they are identical within group A and B respectively which suggests identical sequences also in non-sequenced parts of the fragments in the respective groups. In group C, the compositions of the residual parts are not identical, but fragment V could well be part of fragment VII''. Considering the isolated material not sequenced, i.e.

Table 1
Amino acid compositions, N-terminal amino acids, and Fc-binding activity of protein A fragments

| Amino acid | Protein[a] A | Group A | | Group B | | Group C | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | I | IIA | IIb | III | IV | V | VI'' | VII'' |
| Lys | 52 | 4.0 | 5.2 | 4.1 | 5.3 | 5.8 | 6.8 | 8.5 | 10.5 |
| His | 4 | – | – | 0.9 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 |
| Arg | 4 | 1.1 | 1.0 | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| Asx | 80 | 8.6 | 12.8 | 9.3 | 10.4 | 11.8 | 10.1 | 11.6 | 13.6 |
| Thr | 6 | – | – | – | – | – | 1.0 | 1.0 | 1.0 |
| Ser | 14 | 3.7 | 3.7 | 3.0 | 2.9 | 3.0 | 2.9 | 3.1 | 2.9 |
| Glx | 60 | 11.9 | 12.3 | 11.5 | 10.9 | 11.3 | 13.3 | 14.0 | 15.1 |
| Pro | 28 | 3.0 | 3.1 | 2.9 | 3.3 | 3.2 | 4.3 | 5.3 | 6.3 |
| Gly | 28 | 1.1 | 1.0 | 1.1 | 1.2 | 1.2 | 2.1 | 4.1 | 5.5 |
| Ala | 36 | 4.8 | 5.8 | 5.9 | 5.8 | 6.6 | 4.9 | 4.8 | 5.0 |
| CyS | – | – | – | – | – | – | – | – | – |
| Val | 8 | – | – | – | – | – | 1.0 | 1.0 | 1.0 |
| Met | 3 | 1.0 | 0.9 | – | – | – | – | – | – |
| Ile | 11 | 1.7 | 1.7 | 1.7 | 1.8 | 1.7 | 2.6 | 2.5 | 2.7 |
| Leu | 28 | 5.7 | 6.0 | 7.0 | 6.6 | 6.8 | 5.9 | 5.7 | 6.0 |
| Tyr | 4 | 1.0 | 1.0 | 1.0 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 |
| Phe | 12 | 1.8 | 3.0 | 1.9 | 2.8 | 2.8 | 1.9 | 1.8 | 2.0 |
| Total estimated number of residues | 378 | 51 | 58 | 51 | 54 | 58 | 60 | 68 | 76 |
| Calculated formula weight | 41 500 | 5800 | 6600 | 5900 | 6300 | 6600 | 6800 | 7600 | 8500 |
| N-terminal amino acid | blocked | Glx | Ala | Glx | Phe | Ala | Glx | Glx | Glx |
| Relative Fc-binding Activity[b] | 1 | 445 | 32 | 240 | 2 | 3 | 87 | 63 | 8 |

Experimental details are given in Materials and methods. The calculated formula weights are included.

[a] According to [6].

[b] IgG agglutinates protein A coated sheep red blood cells. The lowest molar amount of fragment which inhibits the hemagglutination activity of added IgG was determined and related to the amount of protein A giving the same effect.

fragments III and VI'', it would be an educated guess to assign them to region B and C respectively as shown in fig.4.

The total molar yield of active material relative to the amount of protein A used for digestion exceeds 200% [9] which was the basis for the statement of at least three active regions. In this work it is shown that at least three active *and* highly homologous regions exist. Of course, further differences may show up, necessitating the subdivision of the fragments into more than three groups. However, if one calculates the molar yields of fragments within each group after purification relative to the amount of protein A used for digestion, they are 47 (A), 46 (B), and 36 (C) percent respectively. This close one to one ratio indicates that only three regions exist, or that a possibly missed region is more easily digested by trypsin into inactive material, or that the active

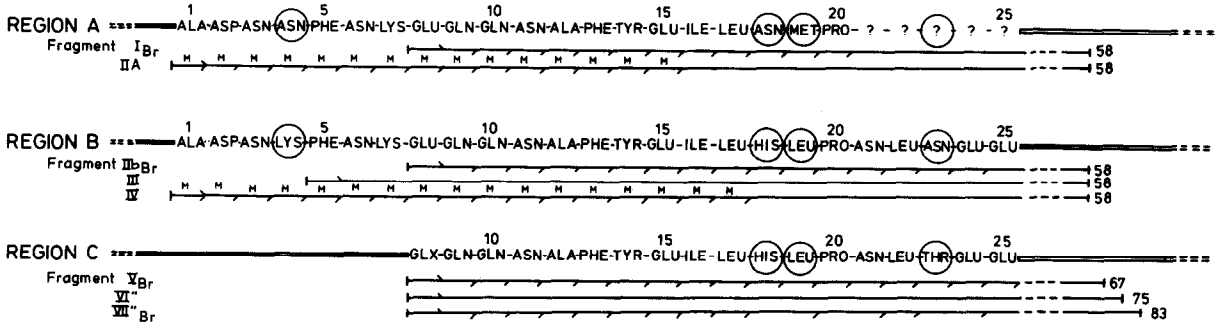


Fig.4. Assignment of the various Fc-binding fragments to different regions within protein A according to amino acid compositions and N-terminal sequences. The data from the groups A–C are lined in order to get maximum homology. The numbering refers to residues within the alignment which is identical to positions in fragments IIA and IV respectively. Assignments based on manual Edman degradations are indicated by (M) and on automatic by (———), on N-terminal determinations by the dansyl method by (———), and on amino acid compositions by (———). Fragments treated with Braunitzer reagent IV are indexed Br. Residues not identical in all regions are circled.

material from this hypothetical region for some reason or other is more difficult to isolated. The involvement of Tyr in the binding reaction [20] and the presence of four Tyr in protein A (6) makes it improbably that more than four regions exist. From the digestion of intact bacteria, only material from region A and B was isolated. This may reflect the location of region C close to a structure obstructing the enzyme in generating fragments from this region, for example the bacterial cell wall.

It is not possible from this study to determine in detail from which amino acids the binding activity arises. However, histidine and methionine clearly are not essential. Comparisons of the relative activity data (table 1) of fragments I and IIA and fragments IIb, III and IV indicate that the N-terminal amino acids are involved, though the data from group C material tell against this theory. Rather, it seems that an amino acid chain of a specific length is required for activity, possibly for stabilization of the structure, or that the loss of possibly active amino acids in the N-terminal part is compensated for in the extended C-terminal parts of fragments V, VI'', and VII''.

Repetitive sequences within a single protein molecule not composed of subunits (at least not linked by S–S bridges) have been reported earlier [21–29], but homologies as high as 80% are rather exceptional. It thus, seems reasonable to suggest that protein A is an evolutionary product of a multiple of an ancestral structural gene. The high conservatism of the system points towards the necessity of having multiple binding sites and well defined structures around the sites in order to preserve the biological functions of the protein. The observed biological activity is the Fc-interaction, though this does not necessarily imply that this is the function that must be preserved.

## Acknowledgements

This investigation was supported by grants from the Medical Faculty, University of Uppsala and from the Swedish Medical Research Council (Proj. no. 13X-2518). The author wishes to express his thanks to Professor J. Sjöquist for stimulating discussions and to Professor B. Blombäck, in whose laboratories the sequencer runs were performed. The technical assistance of Mr T. Möller, Mr J. Tengborn, Mrs S. Söderman and Mrs E. Hjelm is gratefully acknowledged.

## References

[1] Sjöquist, J., Movitz, J., Johansson, I-B. and Hjelm, H. (1972) Eur. J. Biochem. 30, 190–194.
[2] Björk, I., Petersson, B.-Å. and Sjöquist, J. (1972) Eur. J. Biochem. 29, 579–584.
[3] Forsgren, A. and Sjöquist, J. (1966) J. Immunol. 97, 822–827.

[4] Sjöquist, J., Forsgren, A., Gustafson, G. T. and Stålenheim, G. (1967) Cold Spring Harbor Symp. Quant. Biol. 32, 577–581.
[5] Kronvall, G., Gray, H. M. and Williams, R. C., Jr (1970) J. Immunol. 105, 1116–1123.
[6] Sjöquist, J., Meloun, B. and Hjelm, H. (1972) Eur. J. Biochem. 29, 572–578.
[7] Ghetie, V., Nilsson, K. and Sjöquist, J. (1974) Eur. J. Immunol. 4, 500–505.
[8] Biberfeldt, P., Ghetie, V. and Sjöquist, J. (1975) Immunol. Methods 6, 249–259.
[9] Hjelm, H., Sjödahl, J. and Sjöquist, J. (1975) Eur. J. Biochem. 57, 395–403.
[10] Spackman, D. H., Stein, W. H. and Moore, S. (1958) Anal. Chem. 30, 1190–1206.
[11] Gray, W. R. and Smith, J. F. (1970) Anal. Biochem. 33, 36–42.
[12] Hartley, B. S. (1970) Biochem. J. 119, 805–822.
[13] Edman, P. (1970) in: Protein Sequence Determination (Needleman, S. B., ed.) pp. 229–236, Springer-Verlag, New York.
[14] Braunitzer, G., Schrank, B., Ruhfus, A., Petersen, S. and Petersen, U. (1971) Hoppe-Seyler's Z. Physiol. Chem. 356, 1343–1352.
[15] Kulbe, K. D. (1971) Anal. Biochem. 44, 548–558.
[16] Summers, M. R., Smythers, G. W. and Oroszlan, S. (1973) Anal. Biochem. 53, 624–628.
[17] Mendez, E. and Lai, C. Y. (1975) Anal. Biochem. 68, 47–53.
[18] Dawson, R. M. C., Elliot, D. C., Elliot, W. H. and Jones, K. M. (1969) Data for Biochemical Research, p. 530, Oxford University Press, London.
[19] Jepson, J. B. and Smith, I. (1953) Nature (London) 172, 1100–1101.
[20] Sjöholm, I., Bjerkén, A. and Sjöquist, J. (1973) J. Immunol. 110, 1562–1569.
[21] Eck, R. V. and Dayhoff, M. O. (1966) Science 152, 363–366.
[22] Cantor, C. R. C. and Jukes, T. H. (1966) Proc. Nat. Acad. Sci., USA 56, 177–184.
[23] Edelman, G. M. and Gall, W. E. (1969) Ann. Rev. Biochem. 38, 415–466.
[24] Braun, V. and Bosh, V. (1972) Proc. Nat. Acad. Sci., USA 69, 970–974.
[25] Coffee, C. J., Aaronson, R. P. and Frieden, C. (1973) J. Biol. Chem. 248, 1381–1387.
[26] Kula, M. R. (1973) FEBS Lett. 35, 299–302.
[27] Erhan, S. and Greller, L. D. (1974) Int. J. Peptide Protein Res. 6, 175–181.
[28] Koch, G. L. E., Boulanger, Y. and Hartley, B. S. (1974) Nature 249, 316–320.
[29] MacGillivray, R. T. A. and Brew, K. (1975) Science 190, 1306–1307.