Nucleic Acids Research, 1996, Vol. 24, No. 1 157–159

# Protein C deficiency: summary of the 1995 database update

**Pieter H. Reitsma**

Hemostasis and Thrombosis Research Center, Department of Hematology, University Hospital, PO Box 9600, 2333 AA Leiden, The Netherlands

Received October 11, 1995; Accepted October 11, 1995

## ABSTRACT

**The coagulation cascade is controlled by several anticoagulant safeguards that avoid excessive clot formation. Disorders of these anticoagulant mechanisms are an important health problem, as they lead to increased risk of thromboembolism. Protein C deficiency is probably the most extensively studied abnormality in natural anticoagulants. Under the auspices of the Subcommittee on Plasma Coagulation Inhibitors of the Scientific and Standardization Committee of the International Society of Thrombosis and Haemostasis a working party of researchers maintains a database of mutations that have been characterized in the protein C gene. The 1995 update of this database comprises 331 entries that describe 160 unique mutational events. Here essential features of the database are reviewed.**

## PROTEIN C

The key component of the protein C anticoagulant system is protein C. This vitamin K-dependent glycoprotein is synthesized in the liver as a single chain polypeptide and undergoes post-translational modification (β-hydroxylation, γ-carboxylation, proteolytic processing and glycosylation) to give rise to its mature two chain form; the 41 kDa heavy chain and 21 kDa light chains are held together by a disulphide bond. Protein C is an inactive zymogen that is activated by the thrombin/thrombomodulin complex on the endothelial cell surface. Activated protein C exerts its anticoagulant effect (in the presence of its co-factor protein S and phospholipids) by inactivating factors Va and VIIIa and is thought to promote fibrinolysis by neutralizing a plasminogen activator inhibitor (reviewed in 1–3).

There is little doubt about the importance of hereditary protein C deficiency in venous thrombosis. Phenotypically two distinct types of the deficiency state can be recognized: Type I deficiency, the most common, is characterized by a parallel reduction in protein C activity and antigen levels due to the reduced synthesis or stability of normally functioning molecules. In type II deficiency protein C activity is reduced to a greater extent than antigen due to the synthesis of an abnormal protein C molecule exhibiting a reduced specific activity.

Heterozygous clinically 'overt' protein C deficiency is thought to have a prevalence of between 1/16 000 and 1/36 000 in the general population (4,5). However, a much more frequent 'covert' form may also occur in the asymptomatic general population of whom as many as 1/200 may exhibit protein C activity levels consistent with a heritable heterozygous deficiency state (6,7). The absence of thrombosis in many putative heterozygotes appears to argue against plasma protein C concentration being a prime determinant of disease. However, in many individuals/families, heterozygous protein C deficiency is clearly an important independent risk factor for thrombotic disease (8,9) and so other factors must contribute to the likelihood of thrombosis. One important determinent of whether or not a protein C-deficient individual is clinically symptomatic appears to be co-inheritance of the recently described factor V Leiden variant, which is associated with APC resistance (10). This variant occurs at a significantly higher frequency (19%) in individuals with clinically symptomatic protein C deficiency as compared with healthy controls (3%) (10,11).

A large variety of mutations have been shown to underlie hereditary protein C deficiency. To keep track of these mutations a mutation database has been maintained for protein C deficiency, which in its 1995 edition contains 331 entries (160 different) for 315 apparently unrelated probands from 16 different European and American countries (12,13). The majority of these entries are single base pair substitutions that either cause an amino acid replacement (missense), create a premature termination codon (nonsense), disrupt important promoter sequences or alter splice junctions. The remaining entries are almost all insertions or deletions of one or a few nucleotides that either lead to a frame-shift or to amino acid deletions or insertions.

## STRUCTURE OF THE DATABASE

The 1995 update of the database comprises 331 entries which have been ordered according to the nucleotide numbering of Foster *et al.* (14). Each entry is made up of eight columns; a sample of the entries is given in Table 1 (taken from 12). The first column gives the pedigree number. This pedigree number follows the format of each contributing laboratory and varies from entry to entry. Preceding each pedigree number is the telephone code of the country where the proband whose DNA was analysed was born. This is meant to help gain information on the 'epidemiology' of protein C mutations (see below). The second and third column give the plasma protein C antigen and activity data respectively. These are meant to document the distinction between type I and type II deficiency. The fourth column gives the nucleotide position of the mutation and the actual change. In the vast majority of cases these are single nucleotide replacements.

**Table 1.** Sample from the entries in the protein C database of mutations (taken from 12)

| Pedigree | Antigen (% of normal) | Activity (% of normal) | Nucleotide position and mutation | Amino acid change | Co-segregation | Comments | Reference |
|---|---|---|---|---|---|---|---|
| PC-39-I-109 | 16 | 12[c] | 1357, C→T[d] | -11, R→C | 2(3) | Homozygote for type I deficiency mother and father heterozygous | (48) |
| PC-33-070 | 70 | 58[c], 40[a] | 1375, C→T[d] | -5, R→W | 1(4) | Type II | (49) |
| PC-34-003 | 62 | 57[ac], 50[ac] | 1380-86, del 7nt; | -3,-2, del R,K | | Frameshift, stop at codon 16 | (50) |
| PC-33-022 | 60 | 42[a] | 1381, C→T[d] | -3, R→C | 1(1) | | (43) |
| PC-33-033 | 53 | 45[a] | 1381, C→T[d] | -3, R→C | 4(13) | | (43) |
| PC-33-043 | 60 | 63[a] | 1381, C→T[d] | -3, R→C | 2(3) | | (43) |
| PC-34-202 | 52 | 60[c], 33[a] | 1381, C→T[d] | -3, R→C | 4(5) | | (39) |
| PC-33-069 | 150 | 50[a], 130[c] | 1387, C→T[d] | -1, R→C | 3(3) | Type II | (49)(51) |

In some cases, however, small (up to 18 nt long) insertions or deletions are listed. The deleted or inserted nucleotides are then given. The fifth column gives the amino acid change (if any) predicted at the site of mutation. The sixth column indicates whether co-segregation of the mutation with protein C deficiency has been documented. This is important to distinguish between rare 'neutral' polymorphisms and true disease causing mutations. The seventh column leaves room for comments. Examples of the comments given are type of deficiency, whether the proband is a homozygote for protein C deficiency, in the case of a frame-shift mutation the position of the stop codon, etc. The final column gives the reference for the entry.

## FINDINGS FROM AN ANALYSIS OF THE DATABASE

A total of 132 different single base pair substitutions have been noted in the protein C gene. Of these unique mutations no less than 42 (32%) have occurred at CpG dinucleotides and are C→T or G→A transitions, compatible with a model of methylation-mediated deamination (15). As already noted in the first edition of the database, the distribution of CpG site mutations is strikingly non-random in the protein C gene (13). For instance, only one of the 40 entries for exons 4–6 describes a G→A mutation at a CpG dinucleotide. In contrast, the exon 7 entries total 63 mutations, of which 52 are the predicted mutations at CpG dinucleotides (many of these are recurrent). Interestingly, CpG suppression, as measured by the ratio of GpC/CpG frequencies is at a minimum for the region around exons 4–6 (CpG island?) (13,16). This relatively high CpG frequency is both indicative of a lower level of cytosine methylation in the germ line and consistent with the absence of methylation-mediated deamination events.

The fact that for all entries in the database the country of birth of the proband is listed has allowed a study of the spread of protein C gene mutations (17). In this study a strong positive correlation was observed between the initial 'mutational likelihood' of the protein C mutations and dispersal of the gene lesions within and between 16 different countries. In particular, with only a few exceptions, high dispersal was only seen for mutations at the CpG dinucleotides that are known to be hot spots for mutation (see above; 15). The findings suggest that genetic drift and lesion-specific selection have been of minor importance in determining the spectrum of protein C gene mutations and that most multiple reports of particular mutations in different geographical areas reflect recurrent mutation rather than identity-by-descent.

Two research groups have constructed by comparative methods molecular models of the serine protease domain of protein C (17,18). These computer models have been used to study missense mutations causing protein C deficiency. Generally amino acid substitutions causing type I deficiency were predicted to yield energetically unfavourable proteins, since substituting amino acids were predicted to display adverse interactions with neighbouring amino acids. These adverse interactions conceivably lead to disruption of correct protein folding. Substitutions causing type II deficiency were confined to regions of functional significance and were often solvent accessible and therefore were not predicted to interact adversely with neighbouring residues. These models clearly serve as important tools for the understanding of the structure and function of protein C and of the molecular pathology of protein C deficiency.

## DATABASE UPDATE AND AVAILABILITY

The database will continue to be updated on a yearly basis. Novel data may be sent to the author. The database is presently maintained as a WordPerfect file, but will be transferred to a database program

shortly. Copies of the databse on disk are available upon request by email to Reitsma@Rulgca.Leidenuniv.nl.

## ACKNOWLEDGEMENTS

The author wishes to thank Drs Bernardi, Doig, Gandrille, Greengard, Ireland, Krawczak, Lind, Long, Poort, Saito, Sala, Witt and Cooper for their help in maintaining the protein C database of mutations.

## REFERENCES

1 Esmon,C.T. (1994) *Thrombosis Haemostis*, **70**, 29–35.
2 Dahlbäck,B. and Stenflo,J. (1994) In Bloom,A.L., Forbes,C.D., Thomas,D.P. and Tuddenham,E.G.D. (eds), *Haemostasis and Thrombosis.* Chuchill Livingstone, Edinburgh, UK, pp. 671–698.
3 Tuddenham,E.G.D. and Cooper,D.N. (1994) *The Molecular Genetics of Haemostasis and its Inherited Disorders.* Oxford University Press, Oxford, UK.
4 Broekmans,A.W. and Conard,J. (1988) In Bertina,R.M. (ed.), *Protein C and Related Proteins.* Churchill Livingstone, Edinburgh, UK, pp. 160–181.
5 Gladson,C.L., Scharrer,I., Hach,V., Beck,K.H. and Griffin,J.H. (1988) *Thrombosis Haemostis*, **59**, 18–22.
6 Miletich,J., Sherman,L. and Broze,G.,Jr (1987) *New Engl. J. Med.*, **317**, 991–996.
7 Tait,R.C., Walker,I.D., Reitsma,P.H., Islam,S.I.A.M., McCall,F., Poort,S.R., Conkie,J.A. and Bertina,R.M. (1995) *Thrombosis Haemostis*, **73**, 87–93.
8 Bovill,E.G., Bauer,K.A., Dickerman,J.D., Callas,P. and West,B. (1989) *Blood*, **73**, 712–717.
9 Allaart,C.F., Poort,S.R., Rosendaal,F.R., Reitsma,P.H., Bertina,R.M. and Briët,E. (1993) *Lancet*, **341**, 134–138.
10 Koeleman,B.P.C., Reitsma,P.H., Allaart,C.F. and Bertina,R.M. (1994) *Blood*, **84**, 1031–1035.
11 Bertina,R.M., Koeleman,B.P.C., Koster,T., Rosendaal,F.R., Dirven,R.J., De Ronde,H., Van der Velden,P.A. and Reitsma,P.H. (1994) *Nature*, **369**, 74–67.
12 Reitsma,P.H., Bernardi,F., Doig,R.G., Gandrille,S., Greengard,J.S., Ireland,H., Krawczak,M., Lind,B., Long,G.L., Poort,S.R., Saito,H., Sala,N., Witt,I. and Cooper,D.N. (1995) *Thrombosis Haemostis*, **73**, 876–889.
13 Reitsma,P.H., Poort,S.R., Bernardi,F., Gandrille,S., Long,G.L., Sala,N. and Cooper,D.N. (1993) *Thrombosis Haemostis*, **69**, 77–84.
14 Foster,D.C., Yoshitake,S. and Davie,E.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4673–4677.
15 Cooper,D.N. and Youssoufian,H. (1988) *Hum. Genet.*, **78**, 151–155.
16 Bird,A.P. (1986) *Nature*, **321**, 209–213.
17 Krawczak,M., Reitsma,P.H. and Cooper,D.N. (1995) *Hum. Genet.*, **96**, 142–146.
18 Wacey,A.I., Pemberton,S., Cooper,D.N., Kakkar,V.V. and Tuddenham, E.G.D. (1993) *Br. J. Haematol.*, **84**, 290–300.
19 Greengard,J.S., Fisher,C.L., Villoutreix,B., Griffin,J.H. (1994) *Structure Function Genet.*, **18**, 367–380.