Part I Blood Coagulation

1.

The coagulation process: an introduction

Blood coagulation is a field of considerable research interest, so it has already been dealt with in a number of classical reviews (1–3). Recent years have seen a rapid increase in the number of protein sequences determined, by the sequencing of proteins, of cDNA or of genes, with the result that the sequences of all the proteins directly involved in the coagulation system are now known (4) (Table 2). This has led to a considerable increase in our understanding of the mechanism of blood coagulation, although of course merely knowing a sequence does not in itself enable us to understand the components of a system or the way in which they interact.

Traditionally, two pathways of blood coagulation have been distinguished (Fig. 1): the *intrinsic* and the *extrinsic* pathway. These two names allude to the fact that while the intrinsic pathway leads to thrombin formation, using only protein factors that are present in plasma, the extrinsic pathway requires also the presence of a lipid-dependent membrane glycoprotein that does not occur in plasma. However, it is becoming clear that the division of the coagulation process into two branches is somewhat artificial, since several of the components in the two branches appear to interact with one another.

The process of coagulation is classically regarded as a cascade reaction, as originally suggested (5, 6). But recent developments suggest that the cascade concept has outlived its usefulness, because a number of the enzymic reactions leading to the formation of thrombin require the presence of thrombin-activated co-factors if they are to proceed at a physiologically relevant speed. It therefore seems closer to the truth to describe the coagulation system as a cyclic reaction network.

Table 2. *The proteins of the blood coagulation system.*

Protein	Concentration in plasma	Physically-estimated molecular mass (kDa)	Number of amino-acid residues	Calculated molecular mass (kDa)	Gene size (kbp)	Chromosome (and position, if known)
Factor XII	500 nM	80	596	66.9	12	5q33-qter
Plasma prekallikrein	500 nM	85–88	619	69.2	unknown	unknown
Factor XI	35 nM	2 × 75	2 × 607	136	23	4
HMM-kininogen	1 μM	110	626	69.9	27	unknown
Factor IX	80 nM	55–57	415	47	34	Xq27.1
Factor VIII	0.5 nM	>250	2332	264.7	186	Xq27.3
Factor X	170 nM	17 + 42	139 + 306	49.6	27	13q34
Factor V	20 nM	>300	2196	249	unknown	1q21–25
Prothrombin	1.1 μM	72	579	65.7	20.2	11p11–q12
Factor VII	10 nM	50	406	45.5	12.8	13q34
Tissue factor	–	43–53	263	29.6	12.4	1q21–22
Fibrinogen $(A\alpha B\beta\gamma)_2$	6–13 μM	340	2 × 1482	329.6	–	–
$A\alpha$ chain	–	73	610	66.1	unknown	4q26–32
$B\beta$ chain	–	60	461	52.3	10	4q26–32
γ chain	–	53	411	46.4	10.5	4q26–32
Factor XIII (a_2b_2)	70 nM	320	2 × 1372	301	–	–
a-chain	–	80	731	80.5	>160	6p24–25
b-chain	130 nM	80	641	70	unknown	1q31–32.1
Antithrombin III	2.5 μM	54–68	432	49	19	1q23–25
$C\bar{1}$-inhibitor	2 μM	104	478	52.9	17	11p11.2–q13
α_1-proteinase inhibitor	20–50 μM	52	394	44.3	12	14q31–32.3
EPI/LACI[a]	2–3 nM	40	276	32	unknown	unknown
Protein C	60 nM	21 + 41	155 + 262	47.5	11.2	2q14–21
Protein S	150 nM[b]	69	635	70.7	>45	3
Thrombomodulin	–	78	557	58.7	3.7	20
Protein Ca inhibitor	90 nM	57	387	43.8	unknown	unknown

[a] EPI/LACI is an abbreviation for extrinsic-pathway inhibitor, also termed lipoprotein-associated coagulation inhibitor.

[b] Protein S is found partly bound to complement C4b-binding protein, and it is the concentration of free protein S that is given here.

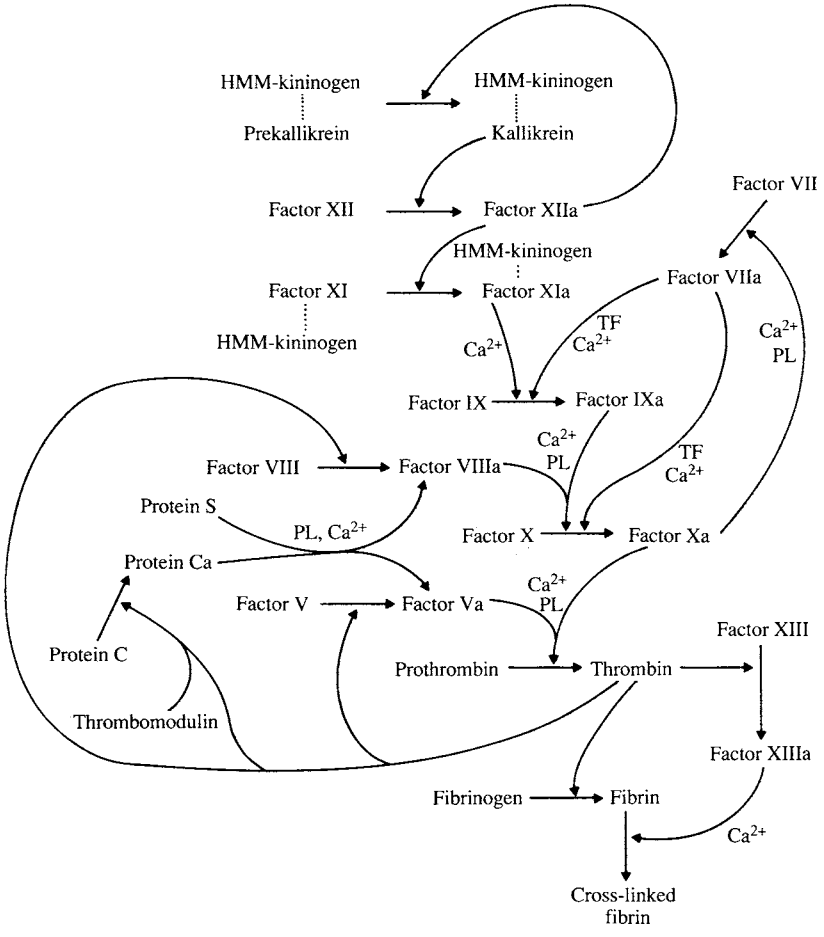

Fig. 1. The blood coagulation system. (TF, tissue factor; PL, anionic phospholipid.)

2. Contact activation

2.1 The proteins of contact activation

Contact activation (reviewed in references 7–13) involves the interaction of four proteins with each other and with a negatively charged surface. The four proteins concerned are factor XII, factor XI, plasma prekallikrein and high-molecular-mass kininogen (HMM-kininogen).

Our knowledge of these four proteins as essential participants in contact activation is due largely to the discovery of individuals with hereditary deficiencies in individual factors. All these deficiencies result in extended coagulation times *in vitro*, although factor XI is the only one of the group whose absence occasionally leads to abnormal coagulation *in vivo*. The discoveries of factor XI (14) and factor XII (15) were the results of genetically determined deficiencies leading to lengthened coagulation times *in vitro*, while plasma kallikrein and HMM-kininogen were already known before studies on plasma lacking them revealed that they are necessary for normal contact activation (16–20).

It was inevitable that some of these coagulation factors should come to bear the names of the patients who were first recognised as lacking them. Factor XII is frequently referred to by its trivial name of Hageman factor (ironically, John Hageman eventually died of a pulmonary embolism (22)). Plasma prekallikrein is called Fletcher factor, while HMM-kininogen is variously called Fitzgerald, Williams, Flaujeac or Reid factor. Factor XI has not been named after a patient, but is occasionally referred to as 'plasma thromboplastin antecedent'.

Before discussing the actual mechanism of contact activation, we shall introduce the individual participant proteins in turn.

2.1.1. *Factor XII*

Factor XII is a single-chained glycoprotein. It circulates in the bloodstream as the inactive precursor (zymogen) of the active serine proteinase factor

Fig. 2. The domain structure of factor XII. → shows the position at which cleavage by plasma kallikrein leads to α-factor XIIa and ⇒ the positions at which extended cleavage by the same enzyme leads to β-factor XIIa. ▼ shows the intron positions, ◆ the N-bound carbohydrate, and ● the O-bound carbohydrate.

XIIa, which takes part in the initial steps of the intrinsic pathway of coagulation. Its molecular mass is estimated from SDS-polyacrylamide gels as 80 kDa, of which some 17% is carbohydrate (21). The concentration of factor XII in normal human plasma is about 40 μg/ml ($\approx$500 nM) (24).

The amino-acid sequence of factor XII has been determined both by protein-sequencing (25, 26) and by cDNA-sequencing (27–29). The structure of the gene for factor XII is also known (30). This gene, located at position q33–qter of chromosome 5 (31), has 14 exons separated by 13 introns and covers about 12 kbp of genomic DNA. (Note: in general the number of introns in a gene is equal to the number of exons minus one. This is the case for all the genes whose structures are mentioned in this book, so the number of introns will not usually be stated explicitly.) The lengths of the exons of the gene for factor XII lie between 57 bp and 320 bp, while the introns are between 80 bp and 4.5 kbp long. Genomic sequencing and cDNA-sequencing have also shown that the signal peptide for factor XII is 19 amino acids in length (see also Chapter 20).

The amino-acid chain of human factor XII (Appendix, Fig. A1) contains 596 amino-acid residues with a calculated molecular mass of 66.9 kDa. If the above figure of 17% carbohydrate is included, then the calculated molecular mass is 80.4 kDa.

The structure of factor XII (Fig. 2) shows a feature also seen in many of the other proteins that we shall meet: a mosaic structure of multiple homology – that is, homology with several other proteins. (The possible evolutionary significance and cause of this are discussed in Chapter 27.) The positions of the introns in the factor XII gene and the relation of these to the amino-acid sequence are shown in Figure 2.

Starting at the N terminus of factor XII, we observe first a domain homologous to the 'type II internally homologous regions' originally found in fibronectin (32, 36), followed by a domain homologous to epidermal growth factor (EGF) and therefore known as the growth-factor domain (33, 34). These are followed in turn by a domain showing homology with

type I homologies of fibronectin (32), another growth-factor domain, a domain homologous to the kringles originally described for prothrombin (35), a portion called the connecting strand, and finally the presumed catalytic region, homologous to known serine proteases.

Although the disulphide pattern of factor XII is not known, the disulphide bridges are assumed to be placed in the same manner as in homologous domains whose disulphide patterns have been determined. This means that the following disulphide pattern in factor XII can be assumed: in the fibronectin type II domain, Cys28–Cys54 and Cys42–Cys69; for the first growth-factor domain, Cys79–Cys91, Cys85–Cys100 and Cys102–Cys111; in the fibronectin type I domain, Cys115–Cys144 and Cys142–Cys151; in the second growth-factor domain, Cys159–Cys170, Cys164–Cys179 and Cys181–Cys190; in the kringle domain, Cys198–Cys276, Cys219–Cys258 and Cys247–Cys271; and in the serine-proteinase domain, Cys340–Cys467, Cys378–Cys394, Cys417–Cys420, Cys386–Cys456, Cys481–Cys550, Cys513–Cys529 and Cys540–Cys571.

The carbohydrate moieties bound to factor XII have been found experimentally at residues Asn230 and Asn414 (N-bound) and at Thr280, Thr286, Ser289, Thr309, Thr310 and Thr318 (O-bound) (25, 26).

The activation of factor XII to factor XIIa occurs by cleavage of the peptide bond between Arg353 and Val354 (26, 27). The resulting derivative of factor XII is termed α-factor XIIa, and it can be converted to β-factor XIIa by cleavage of the bonds Arg334–Asn335 and Arg343–Leu344 (26, 27). All of these three cleavages are performed by plasma kallikrein.

Factor XIIa thus has two chains, consisting of a catalytic serine-proteinase domain disulphide-bonded either to the whole N-terminal region (in α-factor XIIa, residues 1 to 353) or to a nonapeptide (in β-factor XIIa, residues 335 to 343). A covalent linkage between the chains is maintained by a disulphide bond between Cys340 and, as is believed, Cys467. Considerations of homology suggest that the catalytic apparatus in α-factor XIIa involves His40, Asp89 and Ser191 in the processed chain containing the catalytic site, corresponding to His393, Asp 442 and Ser 544 in unprocessed factor XII.

The genetic defect underlying deficiency of factor XII activity has been found in a single case. Here Cys571 was replaced by Ser (37). Since the complete amino-acid sequence of the variant molecule was not determined, it is possible – though unlikely – that other substitutions were also present.

2.1.2. *Plasma prekallikrein*

Plasma prekallikrein is a single-chained glycoprotein that circulates as the (inactive) zymogen precursor of an active serine proteinase. The active protein is plasma kallikrein, and it participates in the contact activation pathway of coagulation. The concentration of plasma prekallikrein in

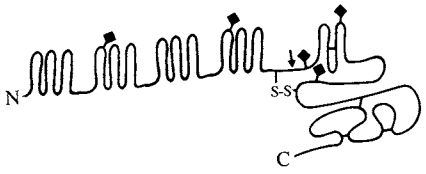

Fig. 3. The domain structure of plasma prekallikrein. → shows the position of cleavage by factor XIIa and ◆ the N-bound carbohydrate.

normal human plasma lies between 40 μg/ml and 55 μg/ml (≈ 500 nM) (38–40). Attempts to isolate it usually yield two forms, with molecular masses estimated from SDS-polyacrylamide gels to be 85 kDa and 88 kDa and with a carbohydrate content of about 15% (41). Isoelectric focussing shows seven bands – that is, plasma prekallikrein appears in seven forms, each of which has its characteristic isoelectric point (42). Four of these forms predominate. Each isoelectric variant contains both molecular-mass variants, but in differing proportions.

The amino-acid sequence of plasma prekallikrein has been determined partially by protein-sequencing (43) and completely by cDNA-sequencing (43) (Appendix, Fig. A2). Included in the cDNA sequence is a signal peptide of 19 amino acids. Neither the gene structure of plasma prekallikrein nor its chromosomal location is known.

The molecular masses calculated from the cDNA sequence are 69.2 kDa without carbohydrate and 80 kDa if the 15% carbohydrate is taken into account. This result is significantly lower than the experimental figures of 85 kDa and 88 kDa (above). However, determinations of molecular mass from SDS-polyacrylamide gels sometimes show deviations, and sedimentation-equilibrium studies indicate a molecular mass of 82 kDa for bovine plasma prekallikrein (44). The heterogeneity in molecular mass is assumed to be due to heterogeneity in the carbohydrate content or in the position of the C terminus, since the purified plasma kallikrein shows no heterogeneity in respect of its N-terminal sequence. Protein-chemical studies have shown that there is N-bound carbohydrate at the residues Asn108, Asn289, Asn377, Asn434 and Asn475.

The structure of plasma kallikrein (Fig. 3) is less complex than that of factor XII. Starting at the N terminus, we observe a succession of four internally homologous domains, of about 90 amino-acid residues each, followed by the serine-proteinase domain. A comparison of the internally homologous regions is given in Figure 4. These domains are almost unique; apart from plasma prekallikrein, they occur in only one other protein, factor XI, to be described below.

The activation of plasma prekallikrein to give plasma kallikrein is carried out by factor XIIa, which cleaves the bond Arg371–Ile372, resulting

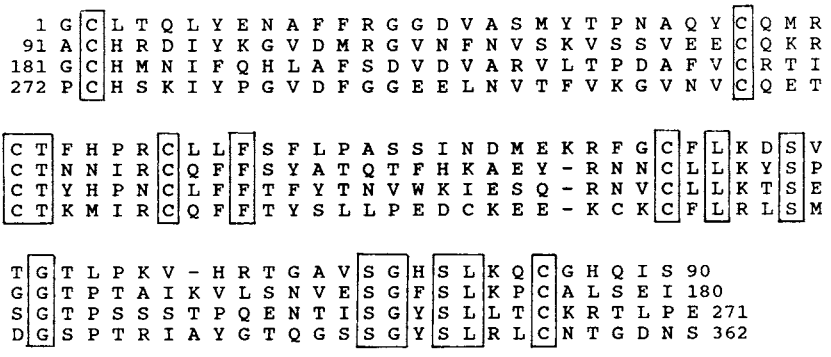

Fig. 4. Internally homologous domains in plasma prekallikrein. Only the positions at which amino acids are the same in all four domains are outlined. Pairwise comparison of the four domains reveals the following degrees of amino-acid identity: domains 1 and 2, 29%; domains 1 and 3, 33%; domains 1 and 4, 26%; domains 2 and 3, 30%; domains 2 and 4, 40%; domains 3 and 4, 25%.

in a heavy chain from the N terminus (52 kDa) and a light chain from the C terminus (33 kDa and 36 kDa) (45). The heterogeneity in molecular mass can thus be attributed to the light chain, which also contains three bound carbohydrate chains. The disulphide-bond pattern has not been demonstrated experimentally, but in the catalytic domain it is assumed to be the same as in other serine proteases, and the position of the disulphide bond between the heavy and light chains can also be deduced from this. Thus the following disulphide bonds are taken to exist: Cys364–Cys484 (which connects the two chains after activation), Cys400–Cys416, Cys498–Cys565, Cys529–Cys544 and Cys555–Cys583. The four internally homologous domains, which together make up the heavy chain of plasma kallikrein, contain six cysteine residues each in homologous positions, and it is expected that these will prove to form three internal S–S connections in each domain. The fourth domain contains two 'extra' cysteines – Cys321 and Cys326 – and these are presumed to make up a fourth internal disulphide bridge in this domain.

Amino-acid-sequencing (43) has revealed a dimorphism at position 124, where the residue that is usually Asn can instead be Ser. On the basis of homology, the catalytic apparatus in plasma kallikrein is taken to involve the residues His44, Asp93 and Ser 188 in the light chain, corresponding to His415, Asp464 and Ser559 in plasma prekallikrein.

2.1.3. *Factor XI*

Factor XI is unique in being a zymogen for a serine proteinase and at the same time consisting of two identical polypeptide chains held together by

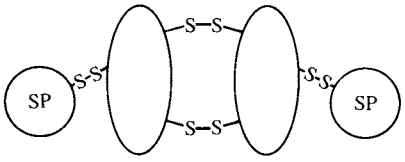

Fig. 5. The two-chain structure of factor XI. (SP, serine-proteinase domain.) The symmetry of the S–S bridged structure has not yet been determined with certainty.

one or more disulphide bridges (Fig. 5). The active enzyme – factor XIa – participates, as do factors XIIa and plasma kallikrein, in the contact activation of blood coagulation. Factor XI is a glycoprotein containing 5% carbohydrate, and its concentration in normal human plasma is about 6 μg/ml ($\approx$35 nM) (46, 47).

The molecular mass of the dimer that makes up the factor XI molecule is estimated by SDS-polyacrylamide gel electrophoresis to lie between 125 kDa and 160 kDa, and the corresponding estimate for the monomer lies between 60 kDa and 83 kDa (48, 49).

The amino-acid sequence of factor XI has been determined largely by cDNA-sequencing (50) (Appendix, Fig. A3); the cDNA sequence includes a signal peptide of 18 residues. The gene for factor XI, found on chromosome 4 (51), covers some 23 kbp of genomic DNA and is distributed among 15 exons of 56 bp to 332 bp that are separated by introns whose sizes range from 88 bp to 4.5 kbp (52).

The molecular mass of the monomer, as deduced from the cDNA sequence, is 68.0 kDa, giving 136.0 kDa for the dimer. The inclusion of 5% carbohydrate brings this value up to about 143 kDa. The carbohydrate is exclusively N-bound, and each monomer possesses five possible sites of N-glycosylation: Asn72, Asn108, Asn335, Asn432 and Asn473. (By 'possible' or 'potential' glycosylation sites we refer here, and at other points in this book, to sites exhibiting the known consensus sequence N-X-S/T for N-glycosylation; see Chapter 19.) The number of N-glycosylation sites used in prekallikrein is also five, but the mass of carbohydrate bound to factor XI is much less than that bound to prekallikrein, which means that either some of the attachment sites in factor XI are not used or else the attached carbohydrate units are much smaller.

The monomeric factor XI chain (Fig. 6) consists of 607 amino-acid residues. It shows homology with prekallikrein over its entire length, with 58% identical amino acids (Fig. 7). In the N-terminal region, the four internally homologous 90–residue domains are found (Fig. 8) and the C-terminal region contains the presumed catalytic serine-proteinase domain. The positions of the introns in the gene for factor XI, relative to the amino-acid sequence, are shown in Figure 6.