pressure vibrational spectroscopic technique on nerve membranes may lead to a better understanding of barotropic phenomena in these tissues, and of the mechanism of anesthesia.

ACKNOWLEDGMENTS

We thank Dr. Yves Geoffrion for assistance in the dissection of the nerves.

REFERENCES

Boulanger, Y., Schreier, S., Leitch, L. C., & Smith, I. C. P. (1980) *Can. J. Biochem. 58*, 986-995.
Boulanger, Y., Schreier, S., & Smith, I. C. P. (1981) *Biochemistry 20*, 6824-6830.
Franks, N. P., & Lieb, W. R. (1982) *Nature (London) 300*, 487-493.
Halsey, M. J., & Wardley-Smith, B. (1975) *Nature (London) 257*, 811-813.
Jaenicke, R. (1983) *Naturwissenschaften 70*, 332-341.
Kelusky, E. C., & Smith, I. C. P. (1984) *Can. J. Biochem. Cell Biol. 62*, 178-184.
Kendig, J. J., & Cohen, E. N. (1977) *Anesthesiology 47*, 6-10.
Lever, M. J., Miller, K. W., Paton, W. D. M., & Smith, E. B. (1971) *Nature (London) 231*, 368-371.
Mao, H. K., Bell, P. M., Xu, J., & Wong, P. T. T. (1982/1983) *Year Book—Carnegie Inst. Wash. No. 82*, 419-421.
Morell, P. (1984) *Myelin*, Plenum, New York.
Mushayakarara, E. C., Wong, P. T. T., & Mantsch, H. H. (1986) *Biochim. Biophys. Acta 857*, 259-264.
Norton, W. T. (1981) *Adv. Neurol. 31*, 93-121.
Norton, W. T. (1984) *Adv. Neurochem. 5*, 47-75.
Pézolet, M., & Georgescauld, D. (1985) *Biophys. J. 47*, 367-372.
Regnard, P. (1887) *C. R. Seances Soc. Biol. Ses Fil. 39*, 406-409.
Roth, S. H. (1979) *Annu. Rev. Pharmacol. Toxicol. 19*, 159-178.
Seelig, A. (1987) *Biochim. Biophys. Acta 899*, 196-204.
Siminovitch, D. J., Wong, P. T. T., & Mantsch, H. H. (1987) *Biophys. J. 51*, 465-473.
Trudell, J. R., Hubbell, W. L., Cohen, E. N., & Kendig, J. J. (1973) *Anesthesiology 38*, 207-211.
Wann, K. T., & Macdonald, A. G. (1980) *Comp. Biochem. Physiol. A 66A*, 1-12.
Wong, P. T. T. (1987a) in *High Pressure Chemistry and Biochemistry* (van Eldik, R., & Jonas, J., Eds.) pp 381-400, D. Reidel, Dordrecht, The Netherlands.
Wong, P. T. T. (1987b) *Vib. Spectra Struct. 16*, 357-445.
Wong, P. T. T., Moffatt, D. J., & Baudais, F. L. (1985) *Appl. Spectrosc. 39*, 733-735.

Articles

Complete cDNA Sequence of Human Complement C1s and Close Physical Linkage of the Homologous Genes C1s and C1r

Mario Tosi,*,‡ Christiane Duponchel,‡ Tommaso Meo,‡ and Cécile Julier§

Unité d'Immunogénétique and INSERM U.276, Institut Pasteur, 75724 Paris Cedex 15, France, and Unité INSERM 91, Hôpital Henri Mondor, 94010 Créteil Cedex, France

Received June 1, 1987; Revised Manuscript Received August 21, 1987

ABSTRACT: Overlapping molecular clones encoding the complement subcomponent C1s were isolated from a human liver cDNA library. The nucleotide sequence reconstructed from these clones spans about 85% of the length of the liver C1s messenger RNAs, which occur in three distinct size classes around 3 kilobases in length. Comparisons with the sequence of C1r, the other enzymatic subcomponent of C1, reveal 40% amino acid identity and conservation of all the cysteine residues. Beside the serine protease domain, the following sequence motifs, previously described in C1r, were also found in C1s: (a) two repeats of the type found in the Ba fragment of complement factor B and in several other complement but also noncomplement proteins, (b) a cysteine-rich segment homologous to the repeats of epidermal growth factor precursor, and (c) a duplicated segment found only in C1r and C1s. Differences in each of these structural motifs provide significant clues for the interpretation of the functional divergence of these interacting serine protease zymogens. Hybridizations of C1r and C1s probes to restriction endonuclease fragments of genomic DNA demonstrate close physical linkage of the corresponding genes. The implications of this finding are discussed with respect to the evolution of *C1r* and *C1s* after their origin by tandem gene duplication and to the previously observed combined hereditary deficiencies of C1r and C1s.

The complement C1 subcomponents C1r and C1s represent a distinct class of serine protease zymogens because of their ability to form a calcium-dependent tetrameric complex (C1s-C1r-C1r-C1s), which interacts with the nonenzymatic subcomponent C1q to yield the ordered C1 structure (Colomb et al., 1984). The activation of C1 is a tightly controlled process triggered by the binding of C1q to immune complexes or to certain nonimmune activators. A fundamental feature of this process is the ability of C1r to autoactivate. As a result, the C1s proenzyme is converted to its active form, which in turn triggers the classical pathway of complement by virtue

‡ Institut Pasteur.
§ Hôpital Henri Mondor.

0006-2960/87/0426-8516$01.50/0

of its highly specific proteolytic activity on C4 and C2 [reviewed in Reid and Porter (1981), Cooper (1985), and Schumaker et al. (1987)].

Structural and biochemical studies (Perkins et al., 1984; Villiers et al., 1985; Weiss et al., 1986; Arlaud et al., 1986) suggested several models for the assembly of C1r, C1s, and C1q (Colomb et al., 1984; Weiss et al., 1986; Schumaker et al., 1986; Arlaud et al., 1987a).

Both C1r and C1s are glycosylated single polypeptide zymogens with molecular masses of about 85 000 daltons that upon activation yield two disulfide-bonded chains. The larger chains of C1r and C1s (A-chains) contain the amino terminus of the corresponding zymogen and are particularly involved in the interactions that lead to formation of the C1s–C1r–C1r–C1s tetramer and probably also in the interactions of the latter with C1q. The shorter chains of C1r and C1s are derived from the carboxy-terminal portion of the zymogens and contain the active site of C1r and C1s, respectively (Sim et al., 1977; Cooper, 1985).

The sequence similarity of portions of C1r and C1s has been documented by comparisons of peptide sequences, confined however to the shorter chain (Carter et al., 1984). Only unlinked peptide sequences, covering about 40% of the A-chain of human C1s, are available for comparisons in this region (Spycher et al., 1986), whereas the sequence of C1r has been completed both at the protein (Arlaud & Gagnon, 1983; Gagnon & Arlaud, 1985; Arlaud et al., 1987b) and at the cDNA level (Leytus et al., 1986a; Journet & Tosi, 1986).

We have recently isolated, from a human liver cDNA library, essentially full-length cDNA clones encoding C1r (Journet & Tosi, 1986) and C1s (Tosi et al., 1985, 1986b). The complete cDNA sequence of C1s, reported in this paper, strongly supports the origin of the *C1r* and *C1s* gene by tandem duplication of a common ancestor and provides clues for understanding, at the structural level, the functional diversification of these related serine protease zymogens.

Experimental Procedures

Screening of a Human Liver cDNA Library. About 30 000 independent colonies from an amplified plasmid library (Tosi et al., 1986a; Journet & Tosi, 1986) were screened by using 23-nucleotide long synthetic probes covering the peptide DWIMKTMQ, located near the carboxyl end of C1s (Carter et al., 1984). The oligonucleotide mixture 5′-T-G-C-A-T-C-G-T-C/T-T-T-C-A-T-G-A-T-C-C-A-G/A-T-C-3′ was synthesized, radioactively labeled, purified, and used as already described (Tosi et al., 1986a), except that the colony filters were hybridized at 50 °C and washed at 53 °C.

DNA Sequence Analyses. Fragments from the inserts of selected C1s cDNA clones (Figure 1) were subcloned into M13 bacteriophage vectors (Messing, 1983) and sequenced by the dideoxy chain termination method of Sanger et al. (1977). The *Bam*HI fragment of clone pHC1s/46, which covers the A-chain of C1s (stippled in Figure 1), was sequenced on both strands by use of a strategy based on the production of oriented deletions (Dale et al., 1985). Briefly, each strand of this fragment was subcloned into M13 bacteriophage vectors. The single-stranded DNA of each subclone was linearized with *Eco*RI after hybridization of a synthetic oligonucleotide complementary to a portion of the polylinker sequence of the vector. Deletions of various length were produced by using T4 DNA polymerase, followed by religation and transformation. In this way overlapping M13 subclones were obtained from which the sequence of each strand could be deduced. Most positions of each strand were sequenced at least twice. The nucleotide sequence corresponding to the B-chain, for which complete protein data are already available (Carter et al., 1984), was also deduced from appropriate M13 subclones of pHC1s/46 after sequencing in the direction and to the extent shown by arrows in Figure 1. Selected portions of additional cDNA clones were sequenced as shown in Figure 1 and described under Results.

RNA Hybridizations. The method of Hagenbüchle et al. (1981) was used to remove from the messengers the poly(A)[1] stretch. Total liver RNA (10 μg) was treated for 15 min at 37 °C with 2 units of RNase H (Genofit, Geneva, Switzerland) in a volume of 50 μL containing 10 mM Tris·HCl, pH 7.4, 130 mM KCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, 2.5 μg of $(dT)_{12-18}$ and 50 units of RNasin (Promega Biotec, Madison, WI). The reaction was stopped by addition of EDTA to 10 mM, followed by two phenol/chloroform extractions. After ethanol precipitation the RNA was electrophoresed alongside 10 μg of untreated liver RNA. The RNA blot methodology was essentially as described (Amor et al., 1985).

Pulsed-Field Gel Electrophoresis. Genomic DNA was prepared from the human lymphoblastoid line L7829, established in the laboratory of Dr. R. White, from a female patient affected by cystic fibrosis. In order to obtain DNA of suitable size, cells were lysed after being embedded in agarose as described (Nakamura et al., 1987). Agarose blocks of a size equivalent to 10 or 15 μg of DNA were treated with restriction endonucleases, and the DNA fragments were separated by orthogonal-field agarose gel electrophoresis (1% agarose) as described (Nakamura et al., 1987). Electrophoresis conditions were 48 h at 13 °C with a constant voltage of 10 V/cm and a pulse time of 40 s.

Results

Sequence of Human C1s. For the isolation of C1s cDNA clones we used a synthetic oligonucleotide probe to screen a library constructed in the *Pst*I site of the plasmid vector pUC9 and representing human liver mRNA fractions size-enriched for C1r and C1s messengers (Tosi et al., 1986a; Journet & Tosi, 1986). The oligonucleotide mixture was designed to correspond to the octapeptide DWIMKTMQ, located near the carboxyl end of human C1s (Carter et al., 1984). About 1 in 1000 clones was detected by this probe in filter hybridizations, and 20 were examined by digestion with *Pst*I. The clone harboring the largest insert, designated pHC1s/46, was subjected to nucleotide sequence analysis, after subcloning of *Bam*HI–*Bam*HI and *Bam*HI–*Hin*dIII fragments (Figure 1). The 2348 bp long insert of pHC1s/46 encodes the entire plasma form of C1s but falls short of comprising the translation initiation codon. On the 3′ side, the insert carries 293 nucleotides of noncoding sequences but does not extend until a polyadenylation site. The full sequence of the pHC1s/46 insert unexpectedly revealed an interruption of the open reading frame at the position marked by an asterisk in Figure 1, suggesting the presence of a cloning artifact in this region or the cloning of an aberrant C1s mRNA molecule. We isolated additional overlapping cDNA clones by rescreening our library with the 5′-most *Bam*HI–*Pvu*II fragment of pHC1s/46 (Figure 1) and sequenced selected areas of two clones, designated pHC1s/5 and pHC1s/22. The former allowed us to extend the C1s mRNA sequence toward the 5′ end, while the latter allowed us to clarify the nature of the anomaly observed

[1] Abbreviations: poly(A), poly(adenylic acid); RNase, ribonuclease; Tris, tris(hydroxymethyl)aminomethane; oligo(dT), oligothymidine; EDTA, ethylenediaminetetraacetic acid; bp, base pair(s); kb, kilobase(s); SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; EGF, epidermal growth factor; LDLR, low-density lipoprotein receptor.

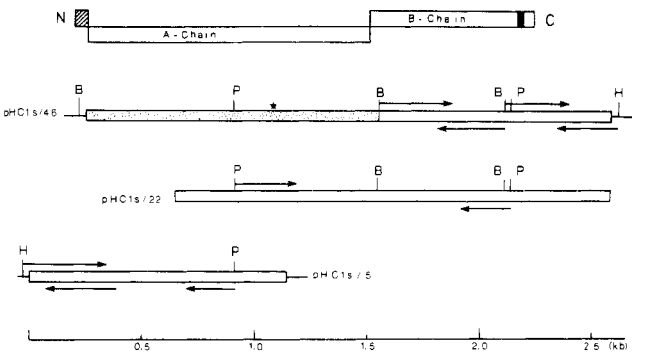

FIGURE 1: Characteristics of C1s cDNA clones and sequencing strategy. The upper diagram represents the C1s precursor protein, consisting of a leader peptide (striped) and of two segments yielding upon activation the A-chain and the B-chain, respectively. The region recognized by the synthetic oligonucleotide probe is marked by a solid rectangle near the carboxyl end. Inserts of three cDNA clones are shown as open bars on which relevant restriction endonuclease sites are indicated (B = *Bam*HI, P = *Pvu*II, H = *Hin*dIII). The stippled portion of clone pHC1s/46 was sequenced on both strands by using a deletion strategy based on single-stranded phage M13 (Dale et al., 1985). Arrows indicate the direction and the extent of sequenced stretches from the same clone, covering the B-chain, and overlaps from other clones. A star marks an interruption of the open reading frame, which is discussed in the text.

in the original cDNA clone. A stretch of seven consecutive adenine residues was found at positions 1058–1064 (Figure 2), one of which had been deleted in clone pHC1s/46 producing the observed frame shift. Indeed, this finding suggests that a cloning artifact is at the origin of the defect. However, the formal possibility that the aberrant cDNA is an exact copy of an authentic C1s mRNA cannot be ruled out without the sequence analysis of a large number of independent cDNA clones or even of the structural genes.

The almost full-length C1s mRNA sequence (2582 nucleotides) shown in Figure 2 reveals a single AUG codon (position 223–225) in phase with the C1s protein sequence. This putative translation initiation codon is preceded by nucleotides that bear little similarity to the consensus sequence found at the translation initiation site of a large number of eucaryotic mRNAs (Kozak, 1984). It is worth noting that the AUG initiation codon of human C1r (Leytus et al., 1986a; Journet & Tosi, 1986) is embedded in a short stretch of nucleotides that display significant similarity (9 out of 10 nucleotides) with the corresponding C1s sequence and thus also deviates from the aforementioned consensus. The putative C1s translation start site defines a 15 amino acid long peptide preceding the N-terminal sequence of the plasma form of human C1s (Sim et al., 1977; Spycher et al., 1986). As expected for a secretory leader peptide (Von Heijne, 1986), this 15 amino acid sequence displays a high content of hydrophobic residues and carries small nonpolar amino acids at positions −3 and −1 preceding the putative signal peptidase cleavage site.

In Figure 2, the peptide bond cleaved upon activation is marked by a arrowhead after the arginine residue 422, according to published data (Sim et al., 1976; Spycher et al., 1986). Note that the *Bam*HI site that delimits on the 3′ side the portion sequenced by use of a deletion strategy (stippled segment in Figure 1) corresponds to residues glycine and serine at positions 426 and 427 (Figure 2). Knowledge of the protein sequence in this portion of the B-chain, as established in two independent studies (Carter et al., 1984; Spycher et al., 1986) rules out the formal possibility that the sequence presented in Figure 2 is short of a small *Bam*HI fragment that might have escaped subcloning and sequencing. For the catalytic B-chain, the protein sequence deduced from cDNA clones confirms the peptide sequence established by Carter et al. (1984), with the possible exception of amino acid residue 630 (Figure 2), where we find a threonine instead of a glycine residue. In the region which upon activation gives rise to the noncatalytic A-chain, the peptides sequenced by Spycher et al. (1986), underlined in Figure 2, were also confirmed, except for amino acid 279. Here we find a cysteine instead of a lysine residue, in better agreement with the conservation of the number and location of cysteine residues between C1s and C1r, as discussed later.

Surprisingly, the A-chain of C1s is only 422 amino acids long, and thus it is significantly shorter than the size predicted from previous analyses of its amino acid composition (Spycher et al., 1986). This discrepancy probably reflects the inaccuracy of size estimates based on amino acid composition and gel filtration analyses of large peptides. The A-chain of C1s is also significantly shorter (422 versus 446 amino acids) than the corresponding region of C1r (Arlaud et al., 1987b; Leytus et al., 1986a; Journet & Tosi, 1986), a finding which accounts for most of the size difference of the two unglycosylated precursor proteins (calculated molecular masses of 76 607 and 79 995 daltons for C1s and C1r, respectively). The smaller size of C1s with respect to C1r represents an unexpected observation, since our previous in vitro translation and immunoprecipitation experiments (Tosi et al., 1986a) had shown that the C1s precursor protein has a slightly lower electrophoretic mobility than the corresponding form of C1r. This discrepancy therefore seems to reflect an anomaly in the relative electrophoretic mobility of these two proteins in SDS–PAGE.

Size Heterogeneity of C1s mRNA. Since the 2582-nucleotide sequence deduced from the overlapping cDNA clones pHC1s/46 and pHC1s/5 does not include a polyadenylated stretch at the 3′ terminus and does not allow precise determination of the 5′ end of the C1s mRNA, we sought an estimate of the size of C1s messengers by Northern blot hybridizations to the human liver RNA from which the cDNA library had been constructed. To this end, we removed the heterogeneous poly(A) stretch from the mRNA population by oligo(dT)-directed RNase H digestion (Hagenbüchle et al., 1981) as shown in Figure 3. At least three classes of C1s mRNA were detected, whose estimated sizes are 3.2, 3.1, and 2.9 kb, respectively. By analogy with the transcriptional heterogeneity of other genes (Tosi et al., 1981; Pfarr et al., 1986) these size differences may be essentially, if not entirely, attributed to variations in the length of the 3′ noncoding region. As reported recently (Leytus et al., 1986a; Journet & Tosi, 1986), the C1r messengers also display size heterogeneity, and in this case multiple polyadenylation signals have been found by cDNA sequencing. As direct evidence is presently lacking in the case of C1s, we cannot rule out the possibility that variations located internally or at the 5′ extremity of the transcripts could also contribute to the C1s mRNA length heterogeneity. In this context, it is interesting to note that Southern blot hybridizations to total genomic DNA yield results consistent with the presence of a single copy of the *C1s* and of the *C1r* gene (data not shown). Alternatively, if there are multiple gene copies at the *C1r* or the *C1s* locus, they must be very similar.

Overall Sequence Homology of C1s and C1r. Previous comparisons of peptide sequences drew attention to the overall similarity of the catalytic B-chains of C1s and C1r (Carter et al., 1984). The complete sequence of C1s allows us to extend such comparisons to the larger, noncatalytic A-chains (Figure 4) and to show that also in this region the two proteins

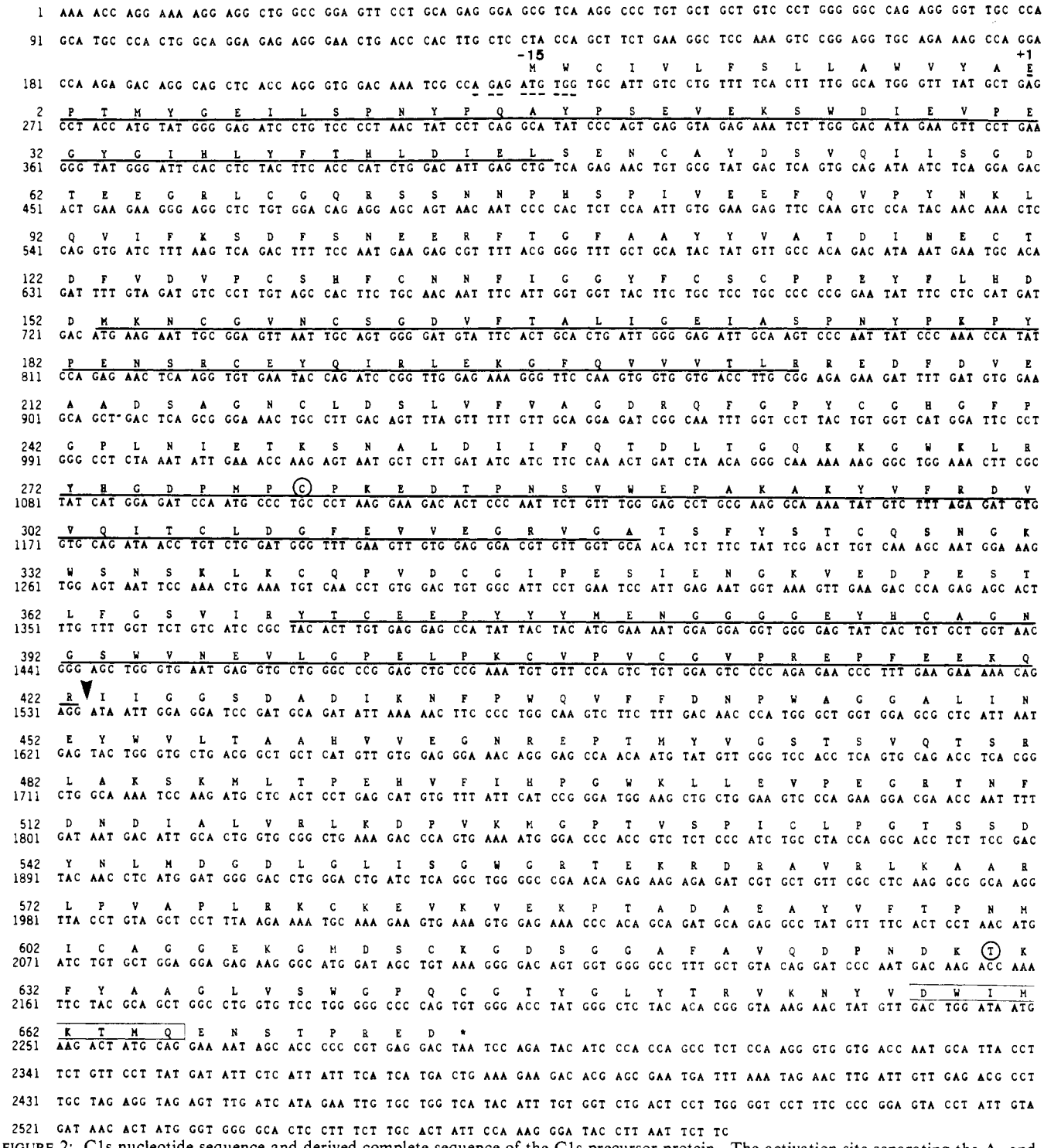

FIGURE 2: C1s nucleotide sequence and derived complete sequence of the C1s precursor protein. The activation site separating the A- and the B-chains is marked by an arrowhead. The carboxy-terminal B-chain differs from its published peptide sequence (Carter et al., 1984) only at position 630 (circled threonine). The boxed octapeptide corresponds to the region recognized by the synthetic probe. Published peptides of the A-chain (Sim et al., 1977; Spycher et al., 1986) are underlined, and the nucleotides conserved with respect to the human C1r cDNA sequence around the putative translation initiation site are underscored.

are highly related in sequence. In particular, cysteine residues are invariant in number and location. Thirty-eight percent of the amino acid residues are identical in the A-chains of C1r and C1s, and thus their sequence similarity is slightly lower than the one previously detected (45%) for the B-chains (Carter et al., 1984). Several short stretches of amino acids appear to be deleted in the A-chain of C1s with respect to the corresponding regions of C1r and account for the smaller size of the A-chain of C1s. However, the composite structure of the A-chain of C1r (Leytus et al., 1986a; Journet & Tosi, 1986; Arlaud et al., 1987b) is mirrored in C1s. Both proteins display a duplicated sequence, marked by solid brackets in Figure 4 and corresponding to domains I and III of C1r, according to the designation introduced by Leytus et al. (1986a). This internal repeat bears no similarity (Leytus et al., 1986a) to known protein sequences and thus appears to be specific to these two related serine proteases. The repeats flank a cysteine-rich stretch of about 40 amino acids (boldface letters in Figure 4), present in both C1s and C1r. The composition of the latter is characteristic of an ubiquitous sequence element found, for example, in the epidermal growth factor and its precursor protein, in several blood clotting proteases,

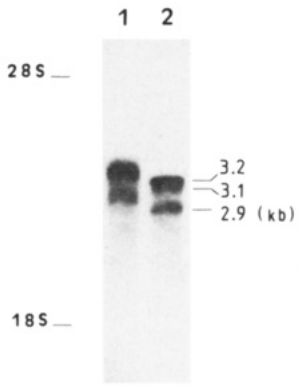

FIGURE 3: Size heterogeneity of C1s mRNA. Lane 1: total human liver RNA (10 μg) was electrophoresed on a 1.5% agarose/formaldehyde gel, transferred to a Gene-Screen membrane (New England Nuclear), and hybridized with the 5′-terminal *Hin*dIII–*Pvu*II fragment of pHC1s/5. Lane 2: a 10-μg aliquot of total RNA was treated before electrophoresis with oligo(dT) and RNase H to remove the heterogeneous poly(A) tail from the mRNAs. This treatment allows the detection of three distinct C1s mRNA species whose size is indicated in kilobases.

```
                I
    SIPIPQKLFGEVTSPLFPKPYPNNFETTTVITVPTGYRVKLVFQQFDLEP
       EP-TMYGEILSPNYPQAYPSEVEKSWDIEVPEGYGIHLYFTHLDIEL

 51 SEGCFYDYVKISADKKSLGRFCGQLGSPLGNPPGKKEFMSQGNKMLLTFH
 47 SENCAYDSVQIISGDTEEGRLCGQRSSNNPHSPIVEEFQVPYNKLQVIFK

                          II     S
101 TDFSNEENGTIMFYKGFLAYYQAVDLDECASRSKLGEEDPQPQCQHLCHN
 97 SDFSNEERFT-----GFAAYYVATDINECTDFVDVP-------CSHFCNN

                                III
151 YVGGYFCSCRPGYELQEDRHSCQAECSSELYTEASGYISSLQYPRSYPPD
135 FIGGYFCSCPPEYFLHDDMKNCGVNCSGDVFTALIGEIASPNYPKPYPEN

201 LRCNYSIRVERGLTLHLKF-LEPFDIDDHQQVH-CPYDQLQIYANGKNIG
185 SRCEYQIRLEKGFQVVVTLRREDFDVEAADSAGNC-LDSLVFVAGDRQFG

                                      IV
249 EFCGKQRPPDL--DTSSNAVDLLFFTDESGDSRGWKLRYTTEIIKCPQPK
234 PYCGHGFPGPLNIETKSNALDIIFQTDLTGQKKGWKLRYHGDPMPCPKED

297 TLDEFTIIQNLQPQYQFRDYFIATCKQGYQLIEGNQVLHSFTAVCQDDGT
284 TPNSVWEPAKAK--YVFRDVVQITCLDGFEVVEGRVGATSFYSTCQSNGK

                V
347 WHRAMPRCKIKDCGQPRNLPNGDFRYTTTMGVNTYKARIQYYCHEPYYKM
332 WSNSKLKCQPVDCGIPESIENGKVEDPE---STLFGSVIRYTCEEPYYYM

397 QTRAGSRESEQGVYTCTAQGIWKNEQKGEKIPRCLPVCGKPVNPVEQRQR 446
379 ENGGG------GEYHCAGNGSWVNEVLGPELPKCVPVCGVPREPFEEKQR 422
```

FIGURE 4: Alignment of the A-chains of C1r (upper line) and C1s (lower line) showing the conservation of amino acid residues. The C1r sequence is from published data (Leytus et al., 1986a; Journet & Tosi, 1986; Arlaud et al., 1987b). The C1s sequence is based on the cDNA data of Figure 2 and on the partial protein sequence of Spycher et al. (1986). The last residue of the A-chains, which precedes the peptide bond cleaved during activation, and the N-linked glycosylation sites (underlined) have been described previously (Arlaud et al., 1987b; Spycher et al., 1986). Note that all the cysteine residues (marked by a star) are conserved. Comparisons of the C1r sequences cited above indicate a polymorphism at position 135 (Arlaud et al., 1987a). Sequence domains are designated with roman numbers, according to Leytus et al. (1986a), and two classes of internal repeats are comprised within solid and interrupted brackets, respectively. Domains with homology to other proteins are shown with boldface lettering.

and in the low-density lipoprotein receptor (LDLR) [see, e.g., Stanley et al. (1986) for a recent compilation of this and other cysteine-rich sequence elements]. It is noteworthy that both C1s and C1r contain a single copy of this element, whereas two, three, and nine repeats are found in blood clotting proteases, in the LDLR protein, and in the EGF precursor, respectively (Doolittle, 1985; Südhof et al., 1985; Doolittle et al., 1984). These cysteine-rich regions of C1s and C1r are aligned in Figure 5A with a consensus derived from the three EGF-like repeats of human LDLR and with the two homologous cysteine-rich elements of human factor X. Interestingly, C1r and C1s differ in length and sequence particularly in the stretch intercepted by the first two cysteines, and a cluster of acidic residues (Glu, Glu, Asp) is present at this site in C1r but not in C1s. Accordingly, the hydropathy profiles differ markedly in this region (G. Arlaud, personal communication). Leytus et al. (1986a) have indeed noticed a sigmoid hydropathy profile in this part of C1r, where a strongly hydrophilic character is followed by a strongly hydrophobic one. In contrast, C1s appears to be essentially hydrophobic in this region.

The carboxy-terminal end of the A-chain of both C1r and C1s consists of two tandemly located domains comprising 60–70 amino acids each and shown with boldface letters in Figure 4. As noted before in the case of C1r (Leytus et al., 1986a; Journet & Tosi, 1986; Arlaud et al., 1987a), these internal repeats resemble a sequence element initially found in the Ba fragment of complement factor B (Morley & Campbell, 1984) and subsequently in the related complement protein C2, in the complement control proteins factor H and C4b-binding protein, and in the complement receptor CR1. Nevertheless, this repeat occurs also in several proteins unrelated to the complement system [see the recent review by Reid et al. (1986) and references cited therein]. We show in Figure 5B an alignment of these repeats of C1r and of C1s (domains IV and V according to the designation introduced previously for C1r) and a comparison with essentially the same consensus sequence proposed by Reid et al. (1986) for this polypeptide family. Within each of the two elements the sequence similarity between C1r and C1s is more pronounced than the conservation of residues between domains IV and V. This observation suggests that both domains were already present in the ancestral gene from which C1r and C1s derived by duplication. Interestingly, at variance with the other sequences of this family, domain V of C1s carries an N-linked glycosylation site (underlined) located within the CxxxGxW stretch, which is an essential component of the consensus derived from these structural elements.

A striking feature of the amino acid sequence of the A-chain of C1s is its high content of negatively charged residues. The distribution of charges along the five sequence elements of the A-chains of C1s and C1r is summarized in Table I. Among these domains the first and the fifth stand out because of their large excess of negatively charged residues in C1s, which is contrasted by the excess of positive charges in the corresponding domains of C1r. These differences in the distribution of charged residues are likely to contribute to the functional diversification of these homologous serine proteases, particularly with respect to their interactions in the C1s–C1r–C1r–C1s tetramer.

Physical Linkage of the C1r and C1s Genes. Since the high degree of sequence similarity of the C1r and C1s proteins suggests that the corresponding genes are related by duplication, we assessed the physical proximity of *C1r* and *C1s* genes in the human genome. Using several common restriction endonucleases, we found by DNA blot hybridization that each gene extends over a stretch of about 15 kb. No cross-hybridization of C1r and C1s sequences was detected (data not shown) even when essentially full-length cDNA probes for either C1r (Journet & Tosi, 1986) or C1s were used (Figure

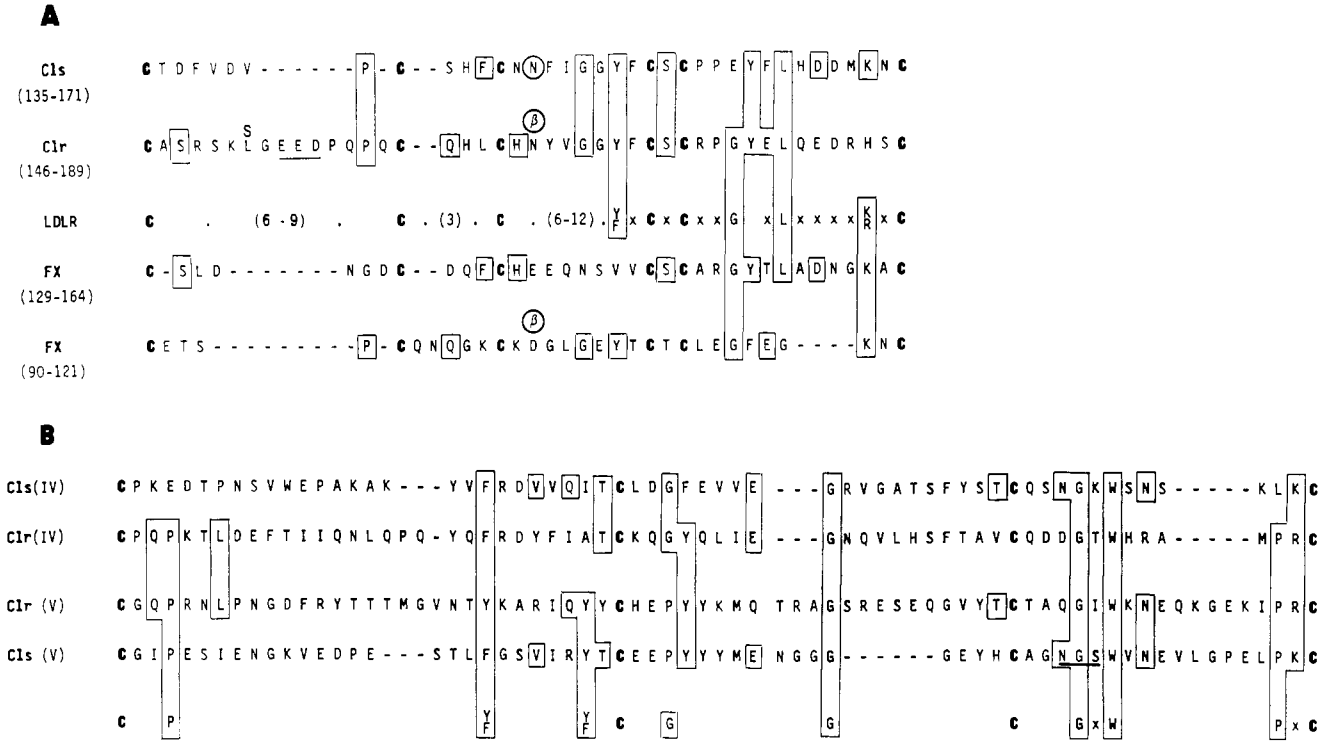

FIGURE 5: Domains of C1r and C1s displaying homology with other proteins. (A) Alignment of the second domains of C1r and C1s with a consensus sequence of the three EGF-like repeats (class B LDLR repeats) of human low-density lipoprotein receptor (Südhof et al., 1985) and with the two homologous repeats of human factor X (Leytus et al., 1986b). The cluster of negatively charged residues Glu-Glu-Asp, a characteristic feature of C1r, is underlined. β indicates sites of posttranslational modifications producing a β-hydroxyasparagine and a β-hydroxyaspartic acid in C1r (Arlaud et al., 1987a) and in factor X (Leytus et al., 1986b; Stenflo et al., 1987), respectively. (B) Alignment of domains IV and V with the consensus sequence (Reid et al., 1986) deduced from comparisons of the complement proteins factor B, C2, C4b-binding protein, factor H, and CR1 and the noncomplement proteins β_2-glycoprotein, interleukin-2 receptor (p55), and factor XIII (β-subunit). Dashes indicate gaps introduced to maximize the sequence similarity. The symbol x is used to indicate that the distance between conserved residues is invariant.

Table I: Distribution of Charged Residues in the Five Sequence Elements of the A-Chains of C1s and C1r

			domain				
	residues	A-chain	I	II	III	IV	V
C1s	R + K	33	5	1	8	8	3
	D + E	64	16	5	14	7	11
	H	8	2	2	1	0	1
	predicted charge[a]	−23 (−31)	−9 (−11)	−2 (−4)	−5 (−6)	+1	−7 (−8)
C1r	R + K	48	10	4	8	5	11
	D + E	53	8	6	13	5	6
	H	10	1	3	3	2	1
	predicted charge[a]	+5 (−5)	+3 (+2)	+1 (−2)	−2 (−5)	+2 (0)	+6 (+5)

[a] A range is given for the expected net charge by assuming that *all* histidines are protonated or that *none* of the histidines is protonated (in parentheses).

1). Therefore, we used restriction endonucleases with very rare recognition sites and separated the resulting DNA fragments by pulsed-field gel electrophoresis in order to detect by hybridization large restriction fragments that might harbor both genes. Figure 6 shows that the *C1s* and the *C1r* genes are both entirely contained within a very large *Not*I fragment (>1000 kb) and also within an *Sfi*I fragment of about 120 kb. The evidence that the latter is indeed a single fragment comprising both genes is provided by the detection of identical hybridization patterns on human DNA completely digested with *Not*I but only partially digested with *Sfi*I (lanes labeled 1–4). Identity of the hybridization on each of these fragments resulting from partial digestion conclusively demonstrates that the *C1r* and *C1s* genes are surrounded by the same *Sfi*I sites and thus lie in close proximity.

Discussion

The complete sequence of human C1s, deduced from overlapping cDNA clones and compared with the published C1r sequences, demonstrates the extensive similarity of the two enzymatic subcomponents of C1, along their entire primary structure. While the 40% identity of amino acid residues supports earlier suggestions (Sim et al., 1977; Cooper, 1985) that these complement serine proteases are derived by gene duplication from a common ancestor, the finding that the *C1r* and *C1s* genes lie in close proximity in the human genome contributes a decisive argument in favor of this notion.

C1s mirrors the composite structure already noted for C1r (Leytus et al., 1986a; Journet & Tosi, 1986; Arlaud et al., 1987b). This finding indicates that the ancestor of the two genes originated by an extension of a serine protease gene, which probably included exons shuffled from other sources. We have presented detailed sequence comparisons for five distinct regions of the noncatalytic A-chain of C1r and C1s, since structural studies have underscored the essential role of this portion of the zymogens in the highly specific interactions

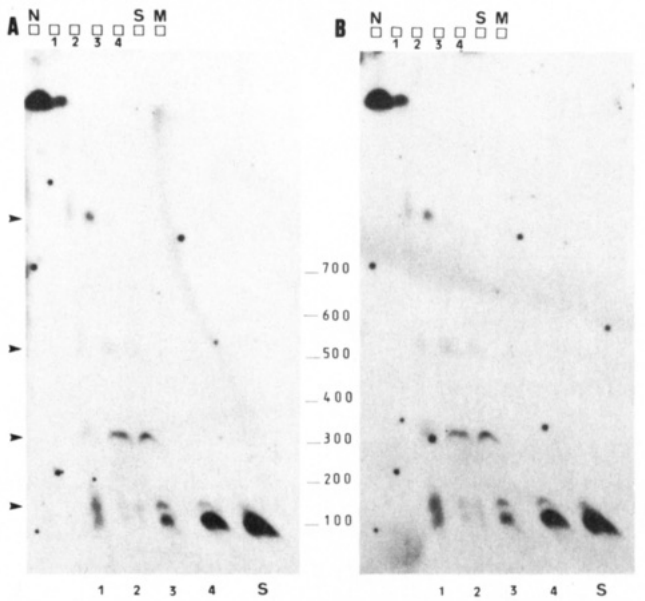

FIGURE 6: Southern blot hybridizations of a C1s probe (panel A) and of a C1r probe (panel B) to large restriction fragments of human DNA separated by pulsed-field agarose gel electrophoresis. Lanes are marked by open squares on top of the autoradiograms and are also identified along the bottom edge by the following symbols: N = restriction endonuclease *Not*I; S = restriction endonuclease *Sfi*I; M = size markers obtained by polymerizing phage λ DNA; 1–4 are tracks carrying human DNA digested with *Not*I to completion and with increasing amounts of *Sfi*I. The filter was first hybridized with the entire plasmid pHC1s/46 (Figure 1), which produces a continuous background on the marker lane. After exposure, this probe was stripped off and the membrane was rehybridized with the entire plasmid pHC1r/253 (Journet & Tosi, 1986). The four hybridization bands resulting from partial digestion with *Sfi*I are indicated by arrowheads. Sizes shown in kilobases between the two autoradiograms were obtained from the ethidium bromide stained phage λ DNA ladder.

within the C1 complex (Colomb et al., 1984; Villiers et al., 1985; Weiss et al., 1986; Shumaker et al., 1986).

Domains I and III are specific structural motifs of C1r and C1s and could be involved in the interaction of the C1s–C1r–C1r–C1s tetramer with C1q, as recently suggested (Arlaud et al., 1987a). Although these domains are homologous, several lines of evidence suggest that they should not be considered as structurally equivalent. For example, Arlaud et al. (1987a) have already pointed out that domain III of C1r contains an additional cysteine residue (position 203 in Figure 4) and an N-linked glycosylation site (position 204) not present in domain I. While the difference in the number of cysteine residues is found also in the corresponding domains of C1s, the comparisons shown in Table I reveal additional differences with regard to the content of charged residues and underscore a particularly striking accumulation of negative charges in the first domain of C1s. These considerations therefore suggest the possibility of different functional roles of domains I and III.

Domain II is a cysteine-rich EGF-like sequence already found in C1r close to its N-terminus (Leytus et al., 1986a; Journet & Tosi, 1986; Arlaud et al., 1987b). As structural studies (Villiers et al., 1985; Weiss et al., 1986) have demonstrated that C1r and C1s monomers interact and bind through their N-terminal sequences, the presence of a similar EGF-like segment also in C1s and the role attributed to this type of domain in other proteins suggest that the interaction of EGF-like domains is an essential feature of the C1r–C1s association. Indeed this type of sequence is found in a variety of secreted or membrane-bound glycoproteins. The latter include the LDL receptor, the *notch* locus of *Drosophila* (Wharton et al., 1985), and the *lin-12* locus of *Caernorhabditis elegans* (Greenwald, 1985). In all cases these domains seem to be exposed to the extracellular environment where they could mediate protein–protein or cell–cell interactions. Moreover, deletion analyses of the LDL receptor (Davis et al., 1987) suggest a role of EGF-like repeats in mediating acid-dependent conformational changes.

It is interesting to note that, at variance with most proteins containing EGF-like sequences, C1r and C1s contain a single domain of this type. Moreover, C1r–C1s binding is calcium-dependent (Cooper, 1985) and an unusual amino acid, β-hydroxyasparagine, has been identified in domain II of C1r (Arlaud et al., 1987a). The latter authors have proposed that, by analogy with similarly modified amino acids in other serum proteins, this residue is involved in calcium binding. Our finding of an asparagine residue at position 134 of the C1s sequence deduced from cDNA studies and the similarity of this portion of the C1s sequence (Figures 4 and 5A) with the consensus proposed for the substrates of β-hydroxylating enzyme(s) (Stenflo et al., 1987) suggest that C1s is also modified posttranslationally at position 134.

A notable difference between C1r and C1s in the EGF-like domain is indicated by a cluster of three negatively charged residues, which is present only in C1r (see Figure 5A) and contributes significantly to the sigmoidal character of the hydropathy profile of this domain, as noted by Leytus et al. (1986a). If, as postulated above, C1r–C1s interactions are mediated by this domain, such a remarkable difference could have important functional implications. However, additional features of the N-terminal sequence of each molecule may also participate in the initiation and stabilization of the C1r–C1s interaction. For example, the striking complementarity of the net charges in domain I of C1s and C1r (see Table I) suggests that ionic bonds may contribute to the interactions involving the first domains.

Domains IV and V of both C1r and C1s are reminiscent of an ubiquitous sequence element present in multiple copies in several other complement components and also in some proteins unrelated to the complement system [reviewed by Reid et al. (1986); see also the recent compilation by Klickstein et al. (1987)]. Domain V of C1s is unusual, because it carries an N-linked glycosylation site within one of the most conserved stretches of this type of sequence (Figure 5B) and also because it displays a large excess of negatively charged residues (Table I). Some of these charges could be involved in the formation of salt bridges with residues of the closely associated B-chain within the "catalytic" domain of each C1s monomer (see Figure 7). In C1r, the catalytic domain has been defined by its resistance to proteolysis (Arlaud et al., 1986), and it comprises the entire B-chain and the portion of the A-chain spanning the fourth and the fifth domain (Arlaud et al., 1987a). Considering the overall sequence similarity and the structural domains of C1s resistant to proteolytic treatments (Villiers et al., 1985; Weiss et al., 1986), it is plausible that the corresponding catalytic domain of C1s has a similar organization.

The apparent complementarity of the net charge between the fifth domains of C1s and C1r may also contribute to the interactions of the C-terminal portions of the four enzymatic subunits within the folded "eight"-shaped conformation of the C1s–C1r–C1r–C1s tetramer, which precedes activation. In the model that represents this state (Arlaud et al., 1986, 1987a), four catalytic domains, each contributed by a C1r or C1s subunit, are brought together in the center of the bell-

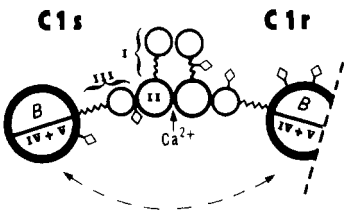

FIGURE 7: Schematic representation of half of the C1s–C1r–C1r–C1s tetramer, showing the calcium-dependent C1s–C1r interactions through the N-terminal domains of the A-chains. The thick dotted line represents the symmetry axis of the C1s–C1r–C1r–C1s tetramer and coincides with the sites of C1r–C1r interaction mediated by the C-terminal catalytic domain of C1r. The scheme extends to the C1s–C1r interaction, the refined model of the C1r dimer proposed by Arlaud et al. (1987a), and emphasizes the postulated role of the second domains in C1s–C1r binding. Folding of the tetramer around C1q arms allows contact of the C1s and C1r catalytic domains (Colomb et al., 1984; Villiers et al., 1985; Weiss et al., 1986; Schumaker et al., 1986; Arlaud et al., 1987a) as indicated by the curved dotted line. Wavy lines within domains I and III are region accessible to proteolytic action (Arlaud et al., 1987a). Note that N-linked carbohydrates differ in number and location between C1s and C1r. Their orientation with respect to the folding of the C1s and C1r subunits is arbitrary. Also note that domains I–V do not necessarily lie within the plane of the figure.

shaped structure provided by C1q. Thus, the four serine esterase regions (B-chains) are surrounded by a total of eight domain IV or domain V elements. The suggestion that this type of sequence, when found in complement components, mediates binding to C3b or C4b (Reid et al., 1986) may also apply to domains IV and V of C1r and C1s (Arlaud et al., 1987a). Indeed, Ziccardi (1986) has documented a novel control mechanism whereby nascent C3b and/or C4b limit(s) the turnover of C1 by inhibiting C1r and C1s activation.

The full-length cDNA clones available for both C1r and C1s, in conjunction with the complete sequence information, will now provide the opportunity to test directly the functional role of the structural features discussed above.

Upon their origin by duplication, the *C1r* and *C1s* genes have apparently followed a peculiar evolutionary path that allowed their functional diversification yet optimized the mutual interactions and maintained the interdependence of their products. It is likely that close linkage has been an important factor in the evolution of these genes. First, it could have played a role in holding together at the population level gene variants whose products were able to produce favorable interactions. Second, the finding that *C1r* and *C1s* are still closely linked could also imply that their coordinate mode of expression depends on a common location within a short chromosomal region. Finally, the persistence of the proximity of the *C1r* and *C1s* loci could merely be the consequence of the very short distance separating the two genes. The last interpretation is borne out by the example of C2 and factor B (Campbell & Bentley, 1985), the other homologous although not interdependent serine proteases of the complement system. Their physical distance is so short that one cannot envisage their separation without dramatically affecting the functional integrity of one or both genes. Our present data, obtained by analyses of the type shown in Figure 6, using additional restriction endonucleases, indicate that *C1r* and *C1s* are entirely contained within a DNA stretch no longer than 50 kb. The details of their structural organization and their precise distance should become apparent by examining appropriate genomic clones.

In all reported cases, hereditary deficiencies of C1r and C1s seem to occur in combination, and this observation led to the suggestion that the corresponding genes are either closely linked or share a common control mechanism (Loos & Heinz, 1986). Indeed, hybridization studies using DNA of somatic cell hybrids and metaphase chromosomes allowed us to show that the *C1r* and *C1s* genes are both located distally on the short arm of human chromosome 12 (Cohen-Haguenauer et al., 1986; Van Cong et al., manuscript in preparation). Moreover, the present demonstration of close physical linkage at the scale of kilobases provides clues to explain the observation of combined hereditary deficiencies. Considering the short distance between *C1r* and *C1s*, it seems likely that in the affected chromosomes a large DNA segment comprising portions of both genes is deleted or else that cis-acting mutations have modified regulatory elements shared by the *C1r* and *C1s* genes.

ACKNOWLEDGMENTS

We thank Gérard Arlaud for numerous valuable discussions at several stages of this work and Gérard Arlaud, Agnès Journet, and Matthieu Lévi-Strauss, who have contributed very useful suggestions to the manuscript. We also thank Christine Verna for skillful editorial assistance.

REFERENCES

Amor, M., Tosi, M., Duponchel, C., Steinmetz, M., & Meo, T. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4453–4457.

Arlaud, G. J., & Gagnon, J. (1983) *Biochemistry 22*, 1758–1764.

Arlaud, G. J., Gagnon, J., Villiers, C. L., & Colomb, M. G. (1986) *Biochemistry 25*, 5177–5182.

Arlaud, G. J., Colomb, M. G., & Gagnon, J. (1987a) *Immunol. Today 8*, 106–111.

Arlaud, G. J., Willis, A. C., & Gagnon, J. (1987b) *Biochem. J. 241*, 711–720.

Campbell, R. D., & Bentley, D. R. (1985) *Immunol. Rev. 87*, 19–37.

Carter, P. E., Dunbar, B., & Fothergill, J. E. (1984) *Philos. Trans. R. Soc. London, B 306*, 293–299.

Cohen-Haguenauer, O., Tosi, M., Meo, T., Van Cong, N., & Frezal, J. (1986) *7th International Congress of Human Genetics*, West Berlin, Sept 22–26, Abstracts, Part II.

Colomb, M. G., Arlaud, G. J., & Villiers, C. L. (1984) *Philos. Trans. R. Soc. London, B 306*, 283–292.

Cooper, N. R. (1985) *Adv. Immunol. 37*, 151–216.

Dale, R. M. K., McClure, B. A., & Houchins, J. P. (1985) *Plasmid 13*, 31–40.

Davis, C. G., Goldstein, J. L., Südhof, T. C., Anderson, R. G. W., Russel, D. W., & Brown, M. S. (1987) *Nature (London) 326*, 760–765.

Doolittle, R. F. (1985) *Trends Biochem. Sci. (Pers. Ed.) 10*, 233–237.

Doolittle, R. F., Feng, D. F., & Johnson, M. S. (1984) *Nature (London) 307*, 558–560.

Gagnon, J., & Arlaud, G. J. (1985) *Biochem. J. 225*, 135–142.

Greenwald, I. (1985) *Cell (Cambridge, Mass.) 43*, 583–590.

Hagenbüchle, O., Tosi, M., Schibler, U., Bovey, R., Wellauer, P. K., & Young, R. A. (1981) *Nature (London) 289*, 643–646.

Journet, A., & Tosi, M. (1986) *Biochem. J. 240*, 783–787.

Klickstein, L. B., Wong, W. W., Smith, J. A., Weiss, J. H., Wilson, J. G., & Fearon, D. T. (1987) *J. Exp. Med. 165*, 1095–1112.

Kozak, M. (1984) *Nucleic Acids Res. 12*, 857–872.

Leytus, S. P., Kurachi, K., Sakariassen, K. S., & Davie, E. W. (1986a) *Biochemistry 25*, 4855–4863.

Leytus, S. P., Foster, D. C., Kurachi, K., & Davie, E. W. (1986b) *Biochemistry 25*, 5098–5102.

Loos, M., & Heinz, H. P. (1986) *Prog. Allergy 39*, 212–231.

Messing, J. (1983) *Methods Enzymol. 101*, 20-78.
Morley, B. J., & Campbell, R. D. (1984) *EMBO J. 3*, 153-157.
Nakamura, Y., Julier, C., Wolff, R., Holm, T., O'Connell, P., Leppert, M., & White, R. (1987) *Nucleic Acids Res. 15*, 2537-2547.
Perkins, S. J., Villiers, C. L., Arlaud, G. J., Boyd, J., Burton, D. R., Colomb, M. G., & Dwek, R. A. (1984) *J. Mol. Biol. 179*, 547-557.
Pfarr, D. S., Rieser, L. A., Woychik, R. P., Rottman, F. M., Rosenberg, M., & Reff, M. E. (1986) *DNA 5*, 115-122.
Reid, K. B. M., & Porter, R. R. (1981) *Annu. Rev. Biochem. 50*, 433-464.
Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T., & Tack, B. F. (1986) *Immunol. Today 7*, 230-234.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Schumaker, V. N., Hanson, D. C., Kilchherr, E., Phillips, M. L., & Poon, P. H. (1986) *Mol. Immunol. 23*, 557-565.
Schumaker, V. N., Zavodszky, P., & Poon, P. H. (1987) *Annu. Rev. Immunol. 5*, 21-42.
Sim, R. B., Porter, R. R., Reid, K. B. M., & Gigli, I. (1977) *Biochem. J. 163*, 219-227.
Spycher, S. E., Nick, H., & Rickli, E. E. (1986) *Eur. J. Biochem. 156*, 49-57.
Stanley, K. K., Page, M., Campbell, A. K., & Luzio, J. P. (1986) *Mol. Immunol. 23*, 451-458.
Stenflo, J., Lundwall, A., & Dahlbäck, B. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 368-372.
Südhof, T. C., Goldstein, J. L., Brown, M. S., & Russell, D. W. (1985) *Science (Washington, D.C.) 228*, 815-822.
Tosi, M., Young, R. A., Hagenbüchle, O., & Schibler, U. (1981) *Nucleic Acids Res. 9*, 2313-2323.
Tosi, M., Journet, A., Colomb, M., & Meo, T. (1985) *Complement 2*, 79.
Tosi, M., Duponchel, C., Bourgarel, P., Colomb, M., & Meo, T. (1986a) *Gene 42*, 265-272.
Tosi, M., Journet, A., Lyonnet, D., Colomb, M., & Meo, T. (1986b) *Protides Biol. Fluids 34*, 453-456.
Villiers, C. L., Arlaud, G. J., & Colomb, M. G. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4477-4481.
Weiss, V., Fauser, C., & Engel, J. (1986) *J. Mol. Biol. 189*, 573-581.
Wharton, K. A., Johansen, K. M., Xu, T., & Artavanis-Tsakonas, S. (1985) *Cell (Cambridge, Mass.) 43*, 567-581.
Von Heijne, G. (1986) *Nucleic Acids Res. 14*, 4683-4690.
Ziccardi, R. J. (1986) *J. Immunol. 136*, 3378-3383.

Investigation of Transition-State Stabilization by Residues Histidine-45 and Threonine-40 in the Tyrosyl-tRNA Synthetase†

Robin J. Leatherbarrow* and Alan R. Fersht*

Department of Chemistry, Imperial College of Science and Technology, London SW7 2AY, U.K.

Received May 14, 1987; Revised Manuscript Received August 13, 1987

ABSTRACT: We have analyzed various mutations involving residues Thr-40 and His-45 in the tyrosyl-tRNA synthetase of *Bacillus stearothermophilus*. The utilization of binding energy in catalysis of tyrosyl adenylate formation from tyrosine and ATP was determined from the free energy profiles for the mutant enzymes. Our results confirm that the side chains of Thr-40 and His-45 provide a binding site for the pyrophosphoryl portion of the transition state of this reaction and for pyrophosphate in the reverse reaction. Deletion of these side chains destabilizes the transition-state by 4.9 and 4.1 kcal mol^{-1}, respectively, consistent with a charged hydrogen-bonding interaction. To examine the role of His-45 further, we constructed the potentially conservative mutations His → Gln-45 and His → Asn-45. Both mutant enzymes are debilitated compared with the native enzyme. The His → Gln-45 enzyme is more active than enzyme in which the complete side chain is deleted (His → Ala-45), and so in this location glutamine is a semiconservative replacement. In contrast, the His → Asn-45 mutation is significantly worse than simple deletion of the side chain, indicating that asparagine at this position causes active destabilization of the transition state compared to His → Ala-45. The amide $-NH_2$ of glutamine may be considered stereochemically equivalent to the ϵ-NH of histidine whereas the amide $-NH_2$ of asparagine is comparable to the δ-NH of histidine. The results suggest that the ϵ-NH rather than the δ-NH group of His-45 is involved in the transition-state stabilization. The large range of effects from "conservative" substitutions at position 45 illustrates the danger of inferring information about binding energies when alternative interactions are introduced by mutation.

The tyrosyl-tRNA synthetase of *Bacillus stearothermophilus* has been the subject of extensive studies using protein engineering [see Leatherbarrow and Fersht (1986) for a recent review]. The experiments have involved systematic mutation of residues around the active site of the enzyme in a manner designed to remove interactions with substrates, products, intermediates, and transition states. Comparison of native and mutant enzymes has allowed us to determine the apparent contribution of various side chains to binding and catalysis (Fersht et al., 1985, 1986a,b).

Tyrosyl-tRNA synthetase catalyzes the formation of enzyme-bound tyrosyl adenylate (Tyr-AMP) from tyrosine (Tyr)

† This work was funded by the MRC of the U.K.

0006-2960/87/0426-8524$01.50/0