Protein Science (1994), *3*:2378-2394. Cambridge University Press. Printed in the USA.

Molecular evolution and domain structure of plasminogen-related growth factors (HGF/SF and HGFl/MSP)

LUIS ENRIQUE DONATE,[1] ERMANNO GHERARDI,[2] N. SRINIVASAN,[1] R. SOWDHAMINI,[1] SAMUEL APARICIO,[2,3] AND TOM L. BLUNDELL[1]

[1] ICRF Structural Molecular Biology Unit, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, United Kingdom
[2] ICRF Cell Interactions Laboratory, Cambridge University Medical School, MRC Centre, Hills Road, Cambridge CB2 2QH, United Kingdom

(RECEIVED July 25, 1994; ACCEPTED September 19, 1994)

Abstract

Plasminogen-related growth factors, a new family of polypeptide growth factors with the basic domain organization and mechanism of activation of the blood proteinase plasminogen, include hepatocyte growth factor/scatter factor (HGF/SF), a potent effector of the growth, movement, and differentiation of epithelia and endothelia, and hepatocyte growth factor-like/macrophage stimulating protein (HGFl/MSP), an effector of macrophage chemotaxis and phagocytosis. Phylogeny of the serine proteinase domains and analysis of intron–exon boundaries and kringle sequences indicate that HGF/SF, HGFl/MSP, plasminogen, and apolipoprotein (a) have evolved from a common ancestral gene that consisted of an N-terminal domain corresponding to plasminogen activation peptide (PAP), 3 copies of the kringle domain, and a serine proteinase domain. Models of the N domains of HGF/SF, HGFl/MSP, and plasminogen, characterized by the presence of 4 conserved Cys residues forming a loop in a loop, have been modeled based on disulfide-bond constraints. There is a distinct pattern of charged and hydrophobic residues in the helix-strand-helix motif proposed for the PAP domain of HGF/SF; these may be important for receptor interaction. Three-dimensional structures of the 4 kringle and the serine proteinase domains of HGF/SF were constructed by comparative modeling using the suite of programs COMPOSER and were energy minimized. Docking of a lysine analogue indicates a putative lysine-binding pocket within kringle 2 (and possibly another in kringle 4). The models suggest a mechanism for the formation of a noncovalent HGF/SF homodimer that may be responsible for the activation of the *Met* receptor. These data provide evidence for the divergent evolution and structural similarity of plasminogen, HGF/SF, and HGFl/MSP, and highlight a new strategy for growth factor evolution, namely the adaptation of a proteolytic enzyme to a role in receptor activation.

Keywords: HGF/SF; HGFl/MSP; kringle; *Met* receptor; plasminogen-related growth factors; serine proteinase

Polypeptide growth factors are a diverse group of proteins that regulate the growth, movement, and differentiation of higher eukaryotic cells and exert their activity through specific membrane receptors that transduce the growth factor signal (Ullrich & Schlessinger, 1990). Cloning and sequencing of a large number of polypeptide growth factors has revealed that these proteins can be grouped in several major families: the insulin/IGF (Humbel, 1990), the EGF/TGFα (Gill et al., 1987; Derynck, 1988), PDGF (Heldin et al., 1993), TGFβ (Massagué et al., 1990), NGF (Ebendal, 1992), FGF (Burgess & Maciag, 1989), and the large family of hematopoietic growth factors (Metcalf, 1989; Arai et al., 1990). Growth factors belonging to these different families do not show significant sequence similarity, although 3 members of different families (PDGF, NGF, and TGFβ) share a common protein fold (Daomin et al., 1992; Murzin & Chothia, 1992; Murray-Rust et al., 1993).

It has been known for many years from sequence data that a number of polypeptide growth factors contain protein domains found in a variety of other proteins with unrelated function. A

Reprint requests to: Tom L. Blundell, ICRF Structural Molecular Biology Unit, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.
[3] Present address: Molecular Genetics Unit, Department of Medicine, Addenbrookes Hospital, Cambridge CB2 2QQ, UK.

Abbreviations: HGFl/MSP, hepatocyte growth factor-like/macrophage stimulating protein; HGF/SF, hepatocyte growth factor/scatter factor; HGF/SF-AP, hepatocyte growth factor/scatter factor activating protein; PAP, plasminogen activation peptide; SCR, structurally conserved region; SVR, structurally variable region; t-PA, tissue-type plasminogen activator; u-PA, urokinase-type plasminogen activator; ACA, ε-aminocaproic acid.

clear example of this is the EGF domain, which is found not only in the EGF/TGFα family of growth factors but also in several enzymes involved in the coagulation, fibrinolytic, and complement cascades, in the matrix protein thrombospondin, and in the low-density lipoprotein receptor (reviewed by Baron & Campbell, 1990). Domains of this type are clearly associated with exon shuffling and duplication (Dorit et al., 1990).

In this paper we deal with a new family of growth factors that illustrate a different and previously unknown evolutionary strategy for growth control in animal cells. The overall domain organization of these molecules is remarkably similar to that of the blood proteinase plasminogen, hence the name plasminogen-related growth factors. There are currently 2 known plasminogen-related growth factors, HGF/SF and HGFl/MSP. HGF/SF was discovered independently as a growth factor for liver cells (hepatocyte growth factor, HGF) (Nakamura et al., 1987, 1989; Miyazawa et al., 1989) and as a fibroblast-derived effector of epithelial movement and interactions (scatter factor, SF) (Stoker et al., 1987; Gherardi et al., 1989; Weidner et al., 1990). The sequence of HGFl/MSP was first reported as a cDNA sharing approximately 50% sequence identity with HGF (hepatocyte growth factor-like, HGFl) (Han et al., 1991), but it has recently been established (Bezerra et al., 1993; Yoshimura et al., 1993) that HGFl is identical to macrophage stimulating protein (MSP), a protein isolated earlier for its activity on macrophage chemotaxis and phagocytosis (Skeel et al., 1991). Because the region of chromosome 3 containing the HGFl/MSP gene is duplicated on chromosome 1 (Welch et al., 1989), it is possible that a complete gene encoding a third plasminogen-related growth factor exists on chromosome 1, but this remains to be established conclusively.

HGF/SF induces growth, movement, and differentiation of target epithelial and endothelial cells through binding to a tyrosine kinase receptor encoded by the *Met* proto-oncogene (Bottaro et al., 1991; Naldini et al., 1991). It is expected that a specific receptor for HGFl/MSP exists on macrophages, and it is possible that one of the 2 *Met*-like kinase receptors recently cloned (Huff et al., 1993; Ronsin et al., 1993) may represent the receptor for HGFl/MSP.

During evolution, plasminogen-related growth factors have lost proteinase activity but have retained the proteolytic mechanism of activation of the proteinases (Naldini et al., 1992; Bezerra et al., 1993; Mars et al., 1993; Miyazawa et al., 1993; Yoshimura et al., 1993). Thus, the activity of these growth factors is controlled not solely by transcription and translation but also through specific proteolytic events that occur outside the cells, and are the same ones that initiate blood clotting and fibrinolysis. HGF/SF and HGFl/MSP therefore provide a link between the coagulation, the fibrinolytic, and the growth-regulation pathways of higher organisms, and there is already substantial evidence that these molecules play a critical role in tissue development, remodeling, and repair.

The discovery of plasminogen-related growth factors poses several interesting questions at both the evolutionary and the structural levels. At the evolutionary level, it is of great interest to understand the genetic events that led to the emergence of molecules with growth regulation activity from proteins that initially had enzyme activity. Equally, at the structural level, it is of interest to establish whether binding of receptors by plasminogen-related growth factors involves the same domains involved in binding substrates by blood proteinases.

In this paper we present 3D models of the hairpin loop of the N (PAP homology) domain, the kringle domains, and the inactive serine proteinase domain of the plasminogen-related growth factors. We address the question of their molecular evolution and discuss a model of dimerization that may mediate receptor activation.

Results

Evolution of plasminogen-related growth factors

HGF/SF, HGFl/MSP, and plasminogen belong to a family of proteins defined by the presence of at least 1 kringle domain, a serine proteinase domain, and an activation domain located between the kringle and the proteinase domains. The other members of this family include apolipoprotein (a), urokinase-type and tissue-type plasminogen activators, prothrombin, factor XII, and the recently cloned HGF/SF activator protein, which has a domain organization essentially identical to that of factor XII (Miyazawa et al., 1993).

Because these proteins contain a variable number of kringles (from 1 to 38) as well as extra domains that may have been acquired or duplicated at different time points, we have reconstructed the evolution of the plasminogen-related growth factors in 2 steps. As a first step, we have calculated the molecular phylogeny of the B chains (the serine proteinase domains) present in a single copy and clearly the earliest common elements of these genes. The results of such analysis are shown in Figure 1 (data calculated from cDNA sequences and branch lengths proportional to the genetic distance). The figure supports the view

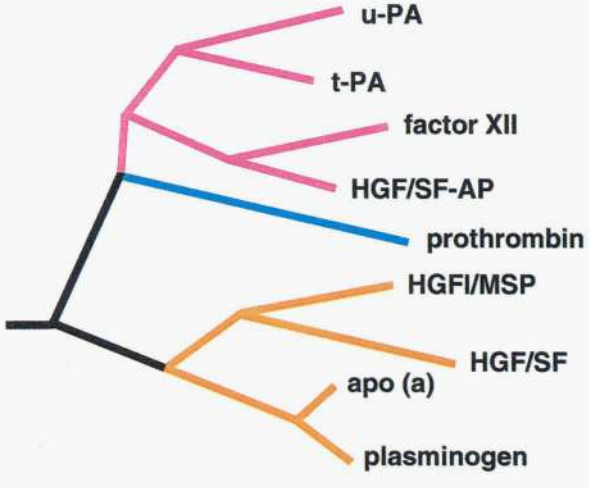

Fig. 1. Phylogeny of the serine proteinase cDNA sequences of proteins containing kringle(s) and a serine proteinase domain. The sequences used were Genbank accessions HUMUKPM (u-PA), HSTPAR (t-PA), HUMCFXII (factor XII), HUMHGFAP (HGF/SF-AP), HSTHR1 (prothrombin), HUMHEPGFA (HGFl/MSP), HUMHHGF (HGF/SF), HSALIPO (apo(a)), and HUMPLASM (plasminogen). The tree was constructed with PHYLIP (Golding & Felsenstein, 1990) using the sequence of human chymotrypsin (Genbank accession HUMCTRP) as an outgroup. The graph shows the results obtained with the Fitch–Margoliash procedure (FITCH). The cladogram has branch lengths proportional to evolutionary distances and was produced by computing a distance matrix with DNADIST and then using DNMAL or FITCH with global rearrangements. The plot was produced with DRAWGRAM.

that the plasminogen-related growth factors evolved along with plasminogen and apo(a) from a common ancestor gene, which duplicated to generate the precursors of the HGF/SF-HGFl/MSP and the plasminogen-apo(a) genes. Both genes duplicated again to generate the 4 genes currently known. The evolution of these 4 genes, therefore, diverged very early from the evolution of the prothrombin, u-PA, t-PA, factor XII, and HGF/SF-AP genes, the latter gene being closely related to the factor XII gene, as inferred from sequence comparison (Miyazawa et al., 1992).

To reconstruct the evolution of the A chains of plasminogen and the 2 plasminogen-related growth factors, we have analyzed their intron-exon boundaries and the sequences of individual kringle domains. The exon boundaries of the HGF/SF, HGFl/MSP, and plasminogen genes are shown in Figure 2 (arrows below and lines across the sequences). Although the length of certain exons differed in the 3 proteins (up to 9 amino acids in the exon encoding the N-terminus of the mature proteins), it is clear that the N domain and kringles 1-3 of HGF/SF, HGFl/MSP, and plasminogen are interrupted by introns in nearly identical positions. The close relation of HGF/SF, HGFl/MSP, and plasminogen kringles 1, 2, and 3 was confirmed by the values of the percentage of sequence identity among them (Table 3A) and by the analysis of their sequences (Fig. 3). Human HGF/SF, HGFl/MSP, and plasminogen kringles 2 and 3 are the only known kringles (of 58 sequenced human kringle domains) that, in addition to the 6 conserved half-cystines at positions 1, 22, 51, 63, 75, and 80 of the kringle domain, contain additional half-cystine residues at positions 4 (kringle 2) and 44 (kringle 3) (the local numbering of kringle sequences is based on that of plasminogen kringle 5).

On the basis of these results and the present domain organization of the genes in the family, we propose that: (1) the gene at the root of the tree in Figure 1 contained a single kringle and serine proteinase domain separated by a region involved in proteolytic activation; and (2) prior to the gene duplication that generated the precursors of the HGF/SF-HGFl/MSP genes and the plasminogen-apo(a) genes, this gene had acquired the exons coding for the N (PAP/PAP homology) domain and 3 copies (kringles 1, 2, and 3) of the kringle domain, probably through initial intragenic duplications of kringle 1.

Kringle 4 of HGF/SF and HGFl/MSP and kringle 5 of plasminogen may have originated from more recent duplications of kringle 1. This interpretation is based on the data in Table 2A and Figure 3 (note the presence of Gly_{60} in all these kringles) as well as the location of the exon boundaries in HGF/SF and HGFl/MSP (Fig. 2). As far as plasminogen kringle 4 is concerned, it is more difficult to establish whether it originated from intragenic duplication or was acquired by exon shuffling.

Analysis of the codons for the 3 residues of the catalytic site—CAC (His), GAT (Asp), and AGT (Ser) in plasminogen; CAG (Gln), GAT (Asp), and TAT (Tyr) in HGF/SF; and CAG (Gln), CAG (Gln), and TAC (Tyr) in HGFl/MSP—suggests that the CAC → CAG (His → Gln) and the AGT → TAT (Ser → Tyr) mutations probably emerged before the separation of the HGF/SF and HGFl/MSP genes and that further mutations in the residues of the catalytic site occurred in the HGFl/MSP gene after divergence from the HGF/SF gene. Thus, prior to the final duplication of the HGF/SF-HGFl/MSP precursor, which produced HGF/SF and HGFl/MSP, the serine proteinase domain of the precursor gene had already lost proteinase activity.

Although the genes for HGFl/MSP and HGF/SF encode proteins of similar length (711 and 728 amino acids, respectively) and have essentially identical intron-exon organization (Fig. 2), the sizes of the 2 genes are remarkably different (4.7 kb and ~70 kb, respectively) due to the presence of much larger introns in the HGF/SF gene (Han et al., 1991; Seki et al., 1991). The GC content of the HGFl/MSP gene is also higher than that of the HGF/SF gene (61.3% vs. 33.7% in the coding sequences). These features indicate that the HGFl/MSP gene more faithfully represents the structure of the parental gene (Doolittle, 1987; Bernardi, 1989) and suggest that the 2 genes have undergone different rates of evolution after duplication.

Models of the N domains of HGF/SF, HGFl/MSP, and plasminogen

To understand the contribution of the different protein domains of HGF/SF (and HGFl/MSP) to receptor binding and biological activity, we have produced models of 3D structures of individual domains: the N-terminal hairpin loop, the kringles, and the serine proteinase.

The N-terminal domain of plasminogen corresponds to the so-called pre-activation peptide, excised by plasmin during plasminogen activation. HGF/SF and HGFl/MSP are the only other proteins that contain a domain homologous to plasminogen PAP. However, the N domain of the plasminogen-related growth factors is not cleaved during activation, and there is evidence from mutagenesis experiments that this domain is essential for the binding of HGF/SF to the *Met* receptor (Matsumoto et al., 1991; Okigaki et al., 1992). It remains to be established whether the N-terminal domain of HGFl/MSP is also essential for binding to the putative HGFl/MSP receptor.

The N domains of plasminogen, HGFl/MSP, and HGF/SF are 73-93 amino acids in length, and all contain a conserved stretch of 23-27 residues defined by 2 disulfide bonds of known connectivity in plasminogen (Wiman, 1973) (Fig. 4A). Deletion of this region alone in HGF/SF is sufficient to abolish receptor binding (Matsumoto et al., 1991).

Although no evidence of homology was discernible, at the sequence level, between the sequences in Figure 4A and other known disulfide-rich systems of known 3D structure, the stereochemistry of this region appears to be highly constrained, and an attempt to model its fold has been based on constraints arising from the covalent crosslinks.

The arrangement of disulfides indicates that this system belongs to the loop within a loop topology (Sowdhamini et al., 1993). When the search for the fold in the disulfide database was performed, the following restrictions were imposed: (1) exact spacing between the first and second Cys residues as observed in the query sequences; and (2) a variability of up to 2 residues in the lengths of the external (1-23) and internal (5-11) disulfide loops.

This search led to a total of 11 potential substructures (listed in Table 1). Although some structural diversity is observed in these segments, four of them (indicated by * in Table 1) have a common motif of the type helix-extended strand-helix. In all of these 4 motifs, the inner disulfide connects a helix and a consecutive extended strand. Interestingly, the natural disulfide A6-A11 in insulin (also a disulfide of similar loop size) has a similar structural feature.

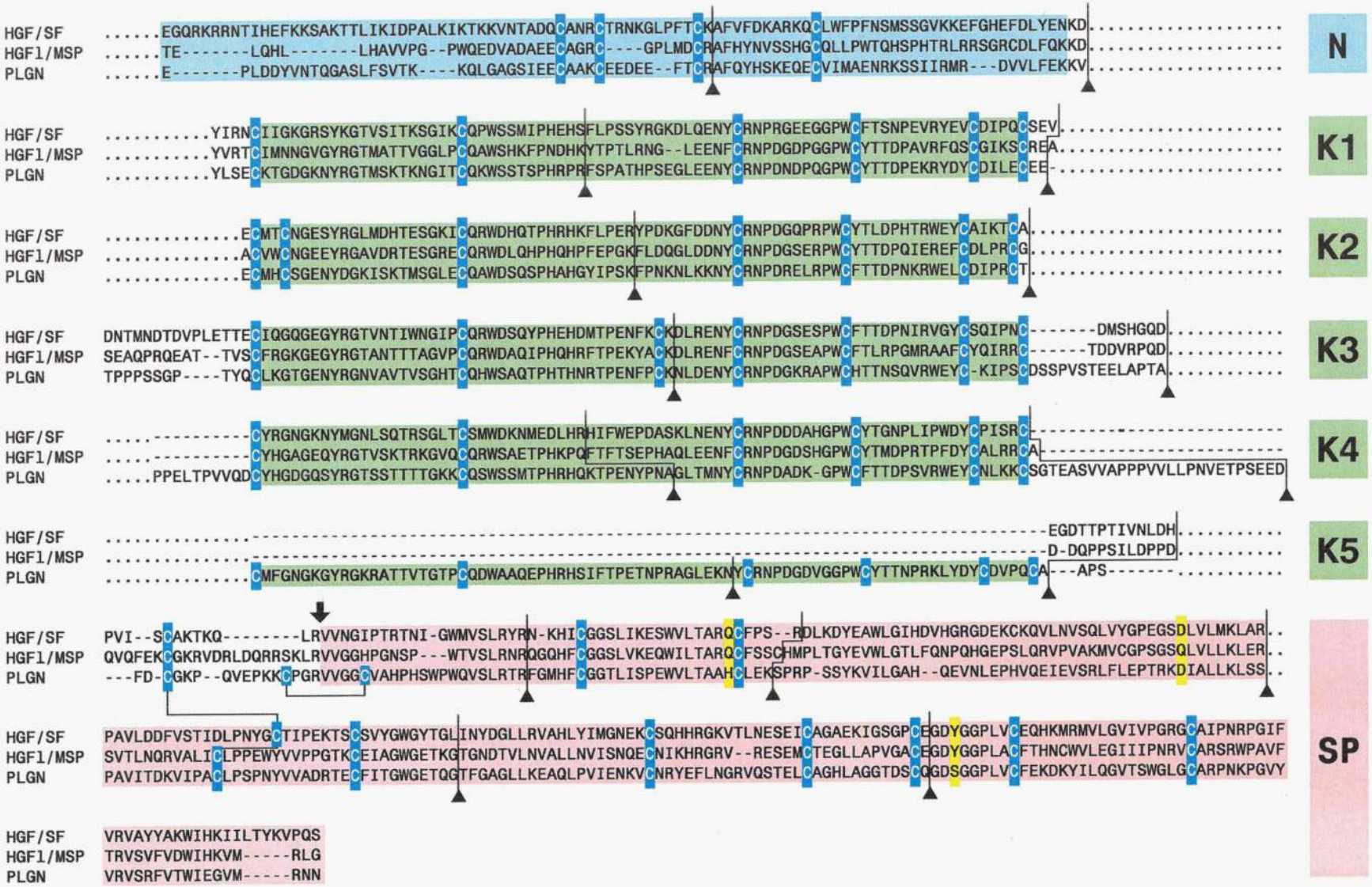

Fig. 2. Protein alignment and exon boundaries of HGF/SF, HGFl/MSP, and plasminogen. The sequences used were Genbank accessions HUMHEPGFA (HGFl/MSP), HUMHHGF (HGF/SF), and HUMPLASM (plasminogen). The alignment was produced using CLUSTAL V and edited manually in order to maximize the alignment in the activation domain. The N domain (light blue) corresponds to the plasminogen activation peptide in plasminogen or the PAP homoiogy domain in HGF/SF and HGFl/MSP. K1–K5 (green) correspond to kringles 1–5 and SP (pink) to the serine proteinase domains. Conserved cysteine are shown as white characters in dark blue background. The 3 residues of the catalytic triad are shown in yellow. Exon boundaries are shown by arrows below the sequences and black lines through the sequences. The arrow above the sequence corresponds to the cleavage site of pro-HGF/SF, pro-HGFl/MSP, and plasminogen.

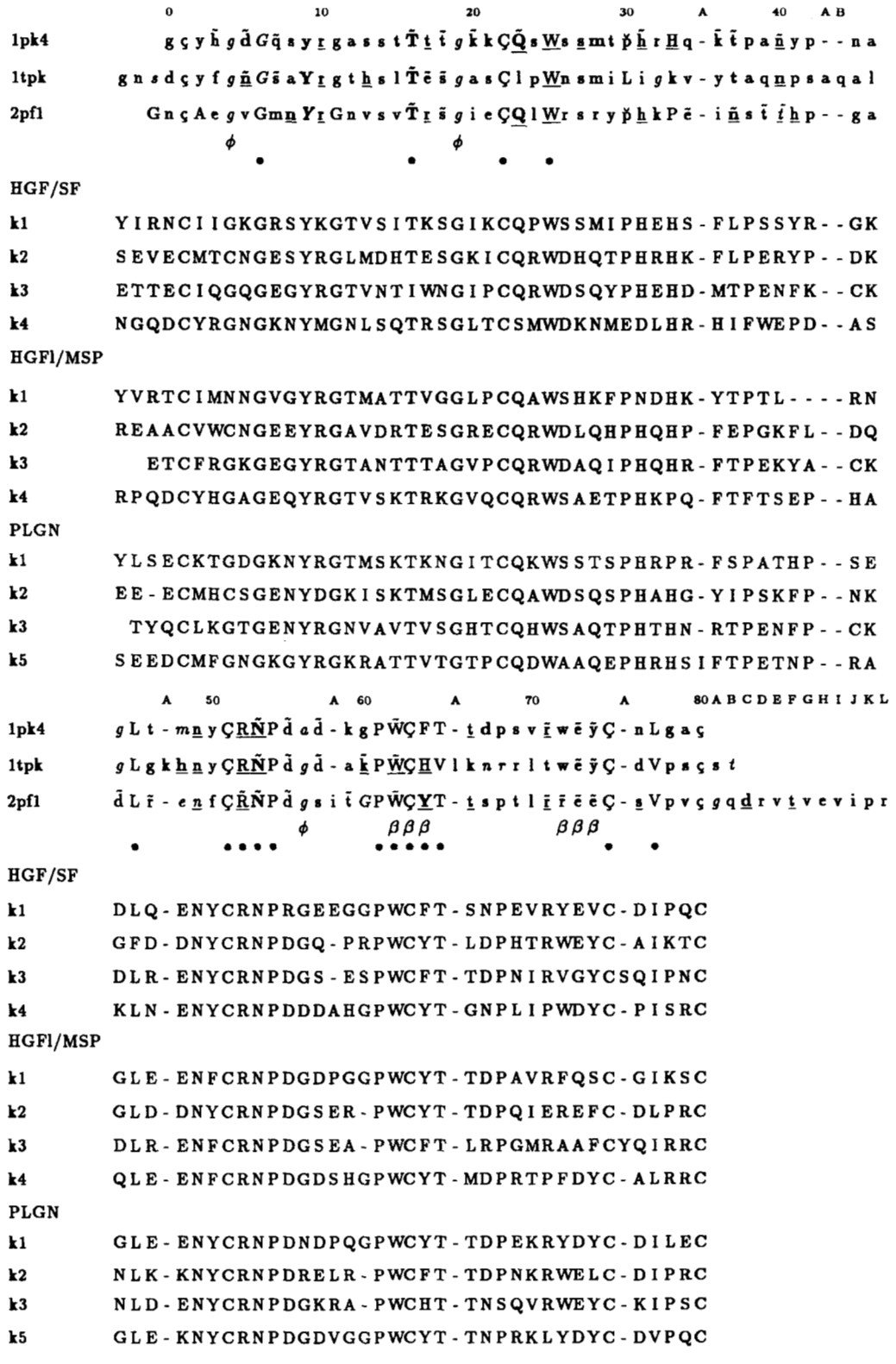

Fig. 3. The structural alignment (JOY alignment) computed for the 3 kringle domains of known structure and the sequence alignments of the kringle domains of HGF/SF, HGFl/MSP, and plasminogen. PDB codes as in Table 2. The standard 1-letter amino acid code is used (C, half cystine; –, a deletion), but with the following additions made to exhibit the similarities and differences in structural environments within the proteins: UPPER CASE, solvent inaccessible; lower case, solvent accessible; *italic*, positive phi torsion angle; *cis*-peptide, breve (i.e., s̆, serine); hydrogen bond to another side chain, tilde (i.e., ñ, asparagine); hydrogen bond to main-chain amide, **bold**; hydrogen bond to main-chain carbonyl, underline; disulfide bond, cedilla; α, α-helical main chain; β, β-strand main chain; ϕ, positive-ϕ main chain; ●, buried conserved residues. Numbering at the top corresponds to that of 1PK4.

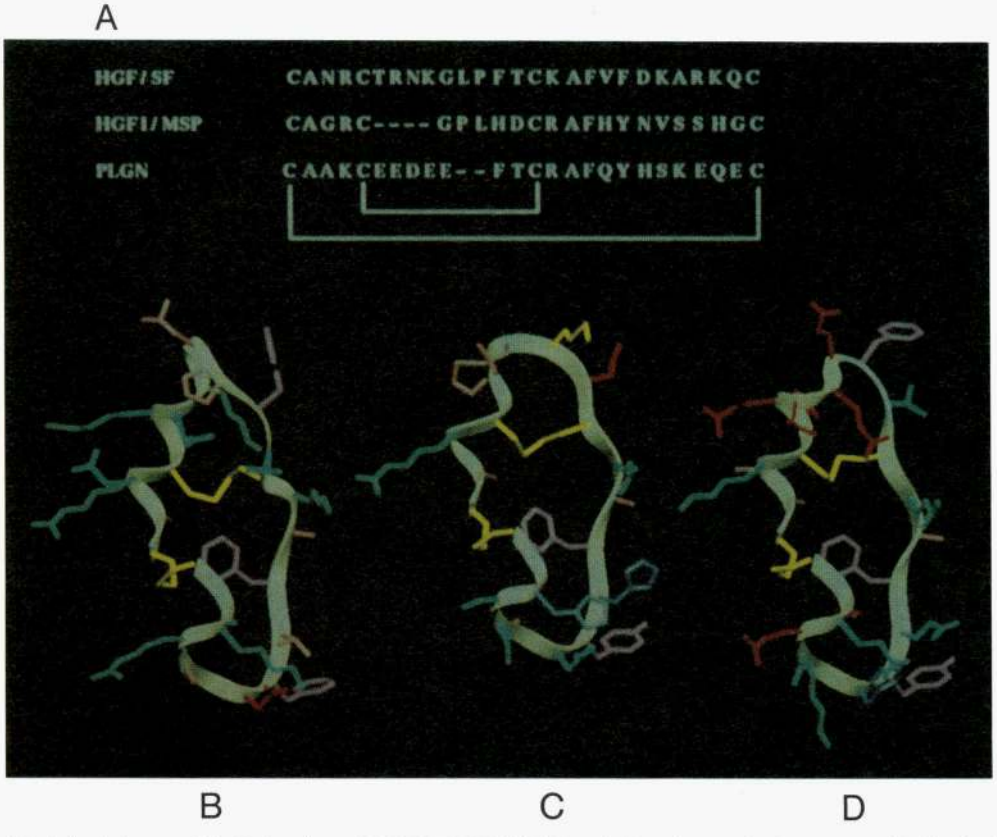

Fig. 4. The hairpin loops of the N domains of HGF/SF, HGFl/MSP, and plasminogen. **A:** Sequences with the disulfide bond connectivity shown in thick lines. The sequences are shown such that the 4 Cys residues are aligned. Residues at positions 2, 12–14, and 16 are also conserved among the 3 proteins (numbering as in HGFl/MSP protein) **B–D:** Energy-minimized 3D models of the hairpin loop of the N-terminal domain of HGF/SF, HGFl/MSP, and plasminogen, respectively. There is a buried phenylalanine at position 12 (shown in pink) interacting with the disulfide cluster (shown in yellow) in all the 3 models.

The N-terminal sequences of the hairpin loops (Fig. 4A) include several charged residues and have good α-helical propensity, whereas the residues immediately following the third Cys have good propensity to lie in an extended conformation. This reinforces the suggestion that the hairpin loop of HGF/SF, HGFl/MSP, and plasminogen has a fold of the type α-loop-β-loop-α, with the first loop of varying length.

Of the 4 sequences mentioned above, one corresponds to the Pro_{183}–Phe_{213} region of citrate synthase (PDB code 1CTS). This segment was used as the initial structure to generate the 3D COMPOSER models of the hairpin loops of the N domain of HGF/SF, HGFl/MSP, and plasminogen, shown in Figure 4B, C, and D after energy minimization.

In all 3 structures, the 2 disulfides have good stereochemistry. The interaction of Phe_{14} with the 4 Cys residues occurs in all the structures; it is well established that such S-π interactions are important stabilizing factors (Reid et al., 1985).

Figure 4B shows that one face of the N- and C-terminal helices of the hairpin loop of HGF/SF is positively charged (3 Arg), whereas the opposite face at the level of the extended strand is highly hydrophobic. Such a prominent positively charged surface is not present in the hairpin loops of HGFl/MSP and plasminogen.

Models of the kringle domains of HGF/SF

The 4 kringle domains of HGF/SF were modeled using the structures of human plasminogen kringle 4 (1PK4) (Mulichak et al., 1991), human tissue-type plasminogen activator kringle 2 (1TPK) (de Vos et al., 1992), and bovine thrombin kringle 1 (2PF1; Seshadri et al., 1991) (Table 2), although the 2PF1 structure was used only to generate the framework and no SCR was modeled on it. As shown in Table 3A, both sequence identity and structural similarity between 1PK4 and 1TPK are quite good (43% and RMS deviation for topologically equivalent Cα atoms is 1.0 Å).

Multiple optimal superposition of the homologous structures resulted in identification of 5 SCRs (although in the case of kringle 3 of HGF/SF, an additional SCR had to be defined; see below). The COMPARER (Sali & Blundell, 1990; Zhu et al., 1992) alignment of these proteins is represented as a JOY alignment (Overington et al., 1990) in Figure 3, in which residue identity, solvent accessibility, side-chain hydrogen bonding, and local conformation are indicated. The amino acid sequence (numbered according to human plasminogen kringle 5, which is 80 residues long) of the kringle domains of HGF/SF is also shown in Figure 3.

Table 1. *Search for substructures with disulfide topology as in the plasminogen PAP domain*[a]

	Residue	*i*	Residue	*j*	S-S loop size	PDB code
SS bond 1	Thr	40	Asn	67	25	2PRK*
SS bond 2	Ala	44	Ala	53	9	2PRK
SS bond 1	Leu	5	Phe	33	26	1ABP
SS bond 2	Val	9	Phe	17	9	1ABP
SS bond 1	Pro	183	Phe	213	29	1CTS*
SS bond 2	Ala	187	Ile	200	13	1CTS
SS bond 1	Pro B	12	Asn	39	27	1TNF
SS bond 2	Val B	16	Trp	28	12	1TNF
SS bond 1	Ala	383	Ala	411	27	3ICD*
SS bond 2	Lys	387	Ala	400	13	3ICD
SS bond 1	Tyr	22	Ala	51	25	1RNH
SS bond 2	Leu	26	Lys	33	7	1RNH
SS bond 1	Glu	216	Asp	244	25	6XIA
SS bond 2	Glu	220	Phe	227	7	6XIA
SS bond 1	Asn	251	Val	277	25	1ACE*
SS bond 2	Asn	255	Glu	261	7	1ACE
SS bond 1	Ala	176	Phe	202	23	1CCP
SS bond 2	Thr	180	Tyr	187	7	1CCP
SS bond 1	His	94	Thr	122	26	1RHD
SS bond 2	Asn	99	Tyr	107	7	1RHD
SS bond 1	Val	95	Val	123	26	1RHD
SS bond 2	Asn	99	Ala	108	8	1RHD

[a] Brookhaven Protein Data Bank codes: 2PRK, proteinase K from *Tritirachium album*; 1ABP, L-arabinose-binding protein from *Escherichia coli*; 1CTS, citrate synthase from pig heart; 1TNF, human tumor necrosis factor alpha; 3ICD, isocitrate dehydrogenase from *E. coli*; 1RNH, selenomethionyl ribonuclease H from *E. coli*; 6XIA, D-xylose isomerase from *Streptomyces albus*; 1ace, acetyl cholinesterase from *Torpedo californica*; 1CCP, yeast cytochrome *c* peroxidase; 1RHD, rhodanase from bovine liver. Asterisks mark examples that correspond grossly to the helix-extended strand-helix motif.

The degree of sequence identity among the kringle domains of HGF/SF is quite high, ranging from 35.4% between kringle 3 and kringle 4 to 50% for the kringle 1–kringle 3 pair (Table 3A). The 4 kringle domains of HGF/SF show higher sequence identity to 1PK4 (from 39.2% for kringle 4 to 50.6% for kringle 1 and 2) and less to 1PTK (from 35.4% for kringle 3 and 4 to 40.2% for kringle 2) (Table 3A).

Kringle 1

Four of the 5 SCRs defined for kringle 1 of HGF/SF were modeled on 1PK4 and one on 1TPK (Table 4). All 6 loops could be modeled on the homologues: four on 1PK4 and one each on 1TPK and 2PF1. The RMS differences between the modeled kringle 1 and 1PK4 and 1TPK are 0.63 Å (for 78 Cα atoms) and 1.21 Å (for 65 Cα atoms), respectively, as shown in Table 3A.

A 5-residue deletion variant of kringle 1 (K1Δ), generated by alternative splicing, that lacks the tract Phe_{36}-Leu_{37}-Pro_{38}-Ser_{39}-Ser_{40} but binds the *Met* receptor with affinity similar to that of full-length kringle 1 (Seki et al., 1990; Rubin et al., 1991; Weidner et al., 1991), was also modeled. All 5 SCRs and 5 of the 6 loops of K1Δ were modeled as for the full-length kringle 1 (see Table 4 for details). The segment 27–43, where the deletion maps, has no corresponding segment in the homologues and was modeled outside the family on 2MEV (Mengo virus coat proteins; Tables 2, 4). The superposed Cα traces of the modeled structures of kringles k1 and K1Δ (Fig. 5) show that the only difference between the 2 structures is in the loop bearing the deletion.

Kringle 2

As in kringle 1, 4 of the 5 SCRs defined by multiple superposition of the homologues for kringle 2 were built on 1PK4 and the other on 1TPK (Table 4). All 6 loops were modeled on the homologues, four on 1PK4 and two on 1TPK (Table 4). The RMS differences between the modeled kringle 2 and 1PK4 and 1TPK (Table 3A) are 0.59 Å (for 79 Cα atoms) and 0.98 Å (for 65 Cα atoms). The kringles have the shape of an oblate ellipsoid

Table 2. *Structures used in the modeling of the kringle domains and serine proteinase domain of HGF-SF*[a]

PDB code	Protein	Source	Resolution (Å)	*R*-factor (%)	Reference for crystal structure used
		Kringle domains			
1PK4	Plasminogen kringle 4	Human	1.9	14.2	Mulichak et al. (1991)
1TPK	Plasminogen activator kringle 2	Human	2.4	18.4	de Vos et al. (1992)
2PF1	Thrombin kringle 1	Porcine	2.5	17.5	Seshadri et al. (1991)
		Serine proteinase domain			
2PTN	Trypsin	Bovine	1.55	19.3	Walker et al. (1982)
3RP2	Mast cell protease	Rat	1.90	19.1	Remington et al. (1988)
3EST	Elastase	Porcine	1.65	16.9	Meyer et al. (1988)
1THR	Thrombin	Human	1.90	15.6	Bode et al. (1992)
		Loops outside homologous families			
2MEV	Coat proteins VP1 + VP2 + VP3 + VP4	Mengo virus	3.0	22.1	Krishnaswamy & Rossmann (1990)
1LAP	Hydrolase	Bovine	2.7	16.9	Burley et al. (1990)
8CAT	Catalase	Bovine	2.5	19.1	Fita & Rossmann (1985)
1PRC	Photosynthetic reaction center	*Rhodopseudomonas viridis*	2.3	19.3	Diesenhofer & Michel (1989)

[a] Atomic coordinates from Brookhaven Protein Data Bank (Bernstein et al., 1977). PDB code, source, resolution, and refinement factor (*R*-factor) of X-ray crystallographic structures are indicated.

Table 3. *Pairwise percent sequence identity and pairwise RMS differences of the modeled kringles and serine proteinase domains of HGF/SF and their corresponding homologues, along with pairwise percent sequence identity among the kringle domains of HGF/SF, HGFl/MSP, and PLGN*[a]

			HGF/SF				HGFl/MSP				PLGN			
	1PK4	1TPK	k1	k2	k3	k4	k1	k2	k3	k4	k1	k2	k3	k4
A.														
1PK4		1.01(65)	0.63(78)	0.51(79)	0.67(77)	0.51(77)								
1TPK	43.0		1.21(65)	0.98(65)	0.99(68)	1.25(77)								
HGF/SF														
k1	50.6	36.6		0.31(81)	0.32(79)	0.32(83)								
k2	50.6	40.2	43.9		0.55(80)	0.45(80)								
k3	48.1	35.4	50.0	47.6		0.45(78)								
k4	39.2	35.4	36.1	36.6	35.4									
HGFl/MSP														
k1			45.7	46.2	47.5	37.0								
k2			42.7	58.0	49.4	36.6	41.2							
k3			50.0	49.4	62.5	37.5	48.7	50.0						
k4			41.0	43.9	42.7	45.8	47.0	43.9	48.7					
PLGN														
k1			50.6	47.6	47.6	43.4	50.6	41.5	42.5	50.6				
k2			48.1	52.5	52.55	40.7	41.8	54.3	45.6	39.5	49.4			
k3			48.1	51.2	56.2	40.7	46.8	48.1	50.0	44.4	45.7	51.2		
k4							48.0	44.9	44.9	51.9	54.4	50.0	51.3	
k5			45.8	50.0	47.5	41.0	53.1	47.6	48.7	49.4	54.2	46.9	50.6	51.9

	2PTN	3EST	3RP2	HGF/SF sp
B.				
2PTN		1.09(207)	1.02(202)	0.90(199)
3EST	40.8		1.15(198)	0.81(209)
3RP2	35.2	35.7		1.03(189)
HGF/SF sp	31.8	34.2	31.7	

[a] Values in the lower triangular matrix represent the sequence identities, and values in the upper triangular matrix represent the RMS differences between the two proteins being compared. Values within parentheses are the number of equivalent Cα atoms. Cα atoms that lie within 2.5 Å of each other in the optimally superposed proteins are considered to be structurally equivalent residues. PDB codes as in Table 2.

comprising very few elements of secondary structure (mainly short stretches of extended β-strands) linked by an extensive net of reverse turns. Features that are characteristic of all the known kringle domains include tight clustering of the sulphur atoms of the 2 inner disulfide bonds, hydrophilic residues exposed on the surface, and hydrophobic residues forming a solvent-inaccessible central core, with the exception of Trp_{72}, which is partially exposed and constitutes a distinctive hydrophobic patch on the kringle surface.

A lysine-binding activity has been described for several kringle domains, and the 3D structures of the lysine-binding pockets of 1PK4 (Wu et al., 1991) and 1TPK (de Vos et al., 1992) have been described. Based on the high degree of conservation in kringle 2 of the residues known to be critical in the lysine-binding pocket of 1PK4, we identify a putative lysine-binding pocket within the structure of HGF/SF kringle 2. Figure 6A shows a general overview of kringle 2, highlighting the side chains of the residues forming the proposed binding pocket. This has the form of an elongated depression on the kringle surface lined by the indole rings of Trp_{62} and Trp_{72}, which form a hydrophobic V-shaped trough. Tyr_{64} and Tyr_{74} make up the remaining 2 walls of the binding site. Some polar/charged residues relevant to lysine binding, such as Asp_{55} and Gln_{57} on the one hand and Arg_{71} on the other hand, are positioned at opposite ends of the trough. Based on the crystal structure of the ε-aminocaproic acid (ACA; in its extended conformation it is a close analogue of lysine) complex of human plasminogen kringle 4, we have modeled the binding of ACA to the putative lysine-binding pocket of HGF/SF kringle 2. A close-up of the binding pocket with the bound ligand is presented in Figure 6B. ACA is docked in such a way that its methylene groups can be accommodated into the hydrophobic environment provided by the side chains of Trp_{62} and Trp_{72}, and its charged termini are within hydrogen bonding distance of several side chains of the binding pocket. ACA can establish 5 hydrogen bonds involving Asp_{55}, Gln_{57}, Tyr_{64}, and Arg_{71} (2 hydrogen bonds) (Fig. 6B).

Kringle 3

To model the segment 75-C of kringle 3, which has an insertion of Ser in position 75a, an SCR additional to those defined by multiple superposition of the homologues was created using residues 78–80 of 1TPK (Table 4). The new loop comprising res-

Table 4. *Selection of fragments for structurally conserved regions (SCRs) and structurally variable regions (SVRs) in the models of the kringles and serine proteinase domains of HGF/SF*

SCRs			SVRs[a]		
SCR in the model[b]	Model based on[c]		SVR in the model[b]	Model based on (beginning from)[c]	
Kringle 1					
2–27	1PK4	2–27	N–2	1TPK	−3
43–46	1PK4	43–46	27–43	1PK4	25
48–57	1TPK	48–57	46–48	1PK4	44
60–65	1PK4	60–65	57–60	2PF1	119
71–7	1PK4	71–75	65–71	1PK4	63
			75–C	1PK4	69
			[d]27–43	2MEV	94[d]
Kringle 2					
2–27	1PK4	2–27	N–2	1TPK	−3
43–46	1PK4	43–46	27–43	1PK4	25
48–57	1TPK	48–57	46–48	1PK4	44
60–65	1PK4	60–65	58–60	1TPK	55
71–75	1PK4	71–75	65–71	1PK4	63
			75–C	1PK4	69
Kringle 3					
2–27	1PK4	2–27	N–2	1TPK	−3
43–46	1PK4	43–46	27–43	1PK4	25
48–57	1TPK	48–57	46–48	1PK4	44
60–65	1PK4	60–65	58–60	1TPK	119
71–75	1PK4	71–75	65–71	1PK4	63
78–80	1TPK	78–80	75–78	1LAP	68
Kringle 4					
2–27	1PK4	2–27	N–2	1TPK	−3
43–46	1TPK	43–4	27–43	1PK4	25
48–57	1TPK	48–57	46–48	1PK4	44
60–65	1PK4	60–65	57–60	2PF1	119
71–75	1PK4	71–75	65–71	1PK4	63
			75–C	1PK4	69
Serine proteinase domain					
16–21	3RP2	16–21	21–26	3EST	149
26–34	2PTN	26–34	34–41	3RP2	198
41–59	2PTN	41–59	59–62	8CAT	334
62–73	3RP2	63–74	73–78	1PRC	165
78–91	3EST	79–91	91–100	3EST	161
100–115	3EST	100–115	115–117	2PTN	113
117–124	2PTN	117–125	124–132	3EST	123
132–145	3EST	132–145	145–151	2PTN	46
151–163	2PTN	151–163	163–179	3EST	161
179–184A	3EST	179–184	184A–18A	3RP2	182
188A–202	2PTN	188A–202	202–203	3EST	200
203–217	3RP2	205–221	217–225	2PTN	215
225–243	3EST	225–243	243–C	1THR	234

[a] N and C refer to N- and C-termini of the polypeptide chain.
[b] For kringles 1–4, residues are numbered according to plasminogen kringle 5. For the serine proteinase domain, residues are numbered according to chymotrypsin.
[c] Protein Data Bank codes as in Table 2.
[d] SVR corresponding to the 5-residue deletion variant of kringle 1. Residues Phe_{35}–Leu_{37}–Pro_{38}–Ser_{39}–Ser_{40} are missing. The rest of the SVRs for this deletion variant are the same as those of the intact kringle 1.

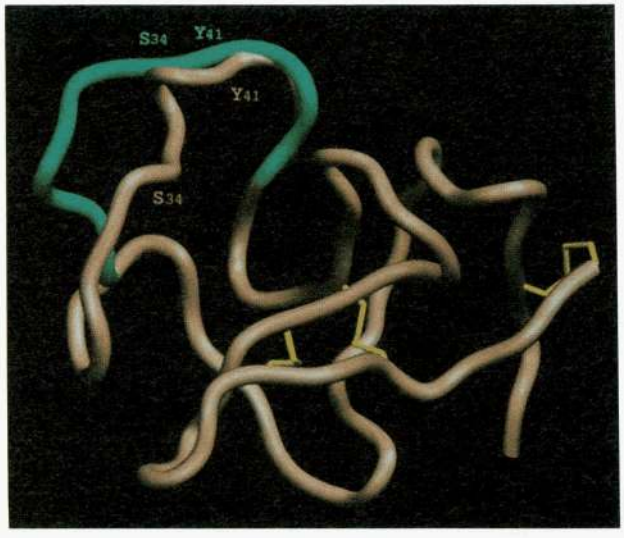

Fig. 5. Solid representation of the Cα traces of energy-minimized models of kringle 1 (shown in pink) and its deletion variant (shown in green) of HGF/SF. The 2 models are completely superimposable between the N-terminus and residue 31 and between residue 43 and the C-terminus. In the modeled deletion variant, the deletion affects the conformation of some residues further away from those that define the boundaries of the deletion, when compared to the model of the intact kringle 1.

idues 75–78 was modeled outside the family on 1LAP (bovine hydrolase; Tables 2, 4). Four SCRs were modeled on 1PK4 and one on 1TPK, whereas two of the SVRs were modeled on 1TPK and three on 1PK4 (Table 4). The RMS differences between the modeled kringle 3 of HGF/SF and the structures of 1PK4 and 1TPK are 0.67 Å (for 77 Cα atoms) and 0.99 Å (for 68 Cα atoms), respectively (Table 3A).

Kringle 4

As shown in Table 4, three of the 5 SCRs defined for kringle 4 of HGF/SF were built on 1PK4 and two on 1TPK. All the 6 segments connecting the SCRs were modeled within the family of homologues, one each on 1TPK and 2PF1 and four on 1PK4. The RMS differences between the modeled kringle 4 and 1PK4 and 1TPK (Table 2A) are 0.51 Å (for 77 Cα atoms) and of 1.25 Å (for 69 Cα atoms), respectively.

Model of the serine proteinase domain of HGF/SF

Of the mammalian serine proteinase structures available in the Brookhaven Protein Data Bank, only those having a sequence identity higher than 30% with the HGF/SF serine proteinase were used in modeling the serine proteinase domain of HGF/SF. Therefore, the serine proteinase domain of HGF/SF was modeled from the structures of bovine trypsin (2PTN) (Walker et al., 1982), rat mast cell proteinase (3RP2) (Remington et al., 1988), and porcine elastase (3EST) (Meyer et al., 1988). As shown in Table 3B, these 3 structures have a good degree of sequence identity (from 35.2% between 3RP2 and 2PTN to 40.8% between 2PTN and 3EST) as well as structural similarity (RMS differences between 1.02 and 1.15 Å).

Multiple optimal superposition of the homologous structures resulted in the identification of 13 SCRs, which were carefully inspected and redefined where necessary to avoid single peptide

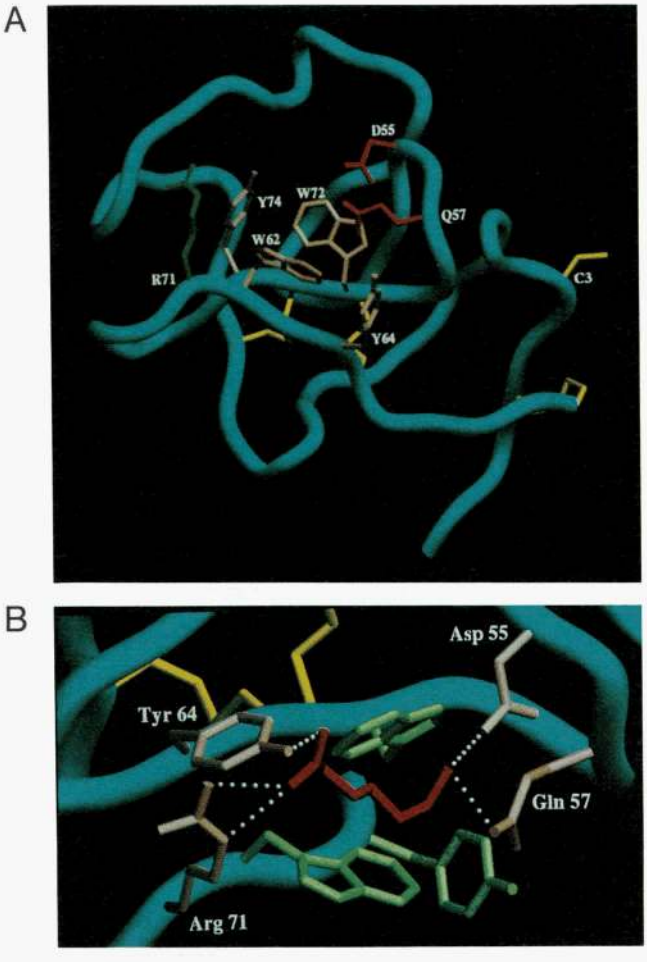

Fig. 6. A: The modeled kringle 2 domain of HGF/SF after the energy minimization. The 3 disulfide bonds and the extra Cys residue are shown in yellow. Aromatic residues forming the V-shaped trough of the putative lysine-binding pocket are shown in magenta, those belonging to the polar/anionic group in red, and that of the cationic group in green. **B:** Close-up of the proposed lysine-binding pocket of modeled kringle 2 of HGF/SF. The docked ACA (a lysine analogue) is shown in red. Full white circles denote the probable hydrogen bonds between the charged moieties of ACA and the residues involved in the pocket (shown in pink).

bonds being considered SVRs. The COMPARER alignment of these structures (represented as a JOY alignment) and the alignment with the sequence of the serine proteinase domain of HGF/SF are shown in Figure 7. The SCRs are also indicated. The sequence identity between the serine proteinase domain of HGF/SF and the homologues ranged from 31.7% to 3RP2 to 34.2% to 3EST (Table 3B).

As indicated in Table 4, of the 13 SCRs defined for the serine proteinase domain of HGF/SF, three were modeled on 3RP2 and five each on 2PTN and 3EST. Of the 13 SVRs, 10 were modeled within the 3 homologous structures (five on 3EST, three on 2PTN, and two on 3RP2), one within the family of mammalian serine proteinases (segment 243-C in the model; built on 1THR, thrombin; Bode et al., 1992) and two of them outside the family. Segment 59–62 was modeled on 8CAT (bovine catalase; Fita & Rossmann, 1985) and segment 73–78 was modeled on 1PRC (photosynthetic reaction center; Deisenhofer & Michel, 1989).

The RMS differences between the model of the serine proteinase domain of HGF/SF and the structures of the homologues are 0.90 Å with 2PTN (for 119 Cα atoms), 0.81 Å with 3EST (for 209 Cα atoms), and 1.03 Å with 3RP2 (for 189 Cα atoms) (Table 3B). In all cases except for 3RP2, the RMS differences between the serine proteinase domain of HGF/SF and the structures used for its modeling were lower than those between the known structures when compared with each other.

The overall fold of the model of the serine proteinase domain of HGF/SF is shown in Figure 8A. The modeled serine proteinase domain resembles very closely the typical bilobal structure of the serine proteinases with known 3D structure. Eight of the 10 Cys residues of the HGF/SF serine proteinase domain are involved in intradomain disulfide bonds (whereas in trypsin, six disulfide bonds are formed, four in elastase, and three in mast cell proteinase); of the other 2 Cys residues (Cys_{79} and Cys_{127} in the model, which correspond to Cys_{561} and Cys_{604} in the HGF/SF sequence, respectively), at least Cys_{127} is involved in interchain interactions (see below). However, the major differences between trypsin and other trypsin-like proteinase and the HGF/SF serine proteinase domain concern the catalytic triad and the S1 specificity pocket. As shown in Figure 8A, 2 of the 3 catalytic residues of trypsin (His_{57}, Asp_{102}, and Ser_{195}) are replaced in the HGF/SF serine proteinase; His_{57} and Ser_{195} have been substituted by Gln (Gln_{534} in the HGF/SF sequence) and Tyr (Tyr_{673} in the HGF/SF sequence) residues, respectively. Figure 8B shows a close-up of both the catalytic site and the residues forming the putative S1 specificity pocket of the HGF/SF serine proteinase domain. In the latter, Asp_{189} and Ser_{190}, the 2 residues located at the bottom of the S1 pocket in trypsin, which are critical for arginine specificity, have been replaced by Gly (Gly_{667} in HGF/SF sequence) and Pro (Pro_{668} in HGF/SF sequence) residues, respectively. Also in the S1 pocket, Ser_{217} (which lies on one of the walls of the pocket) has been replaced by an Arg (Arg_{647} in HGF/SF sequence) residue; other less drastic substitutions are of Gln_{192} by Glu (Glu_{670} in HGF/SF sequence), Val_{227} by Ile (Ile_{705} in HGF/SF sequence), and Tyr_{228} by Phe (Phe_{706} in HGF/SF sequence).

Discussion

This analysis indicates that the HGF/SF and HGFl/MSP genes have evolved along with the plasminogen and apolipoprotein (a) genes from a common ancestor gene. If a constant rate of mutation of 0.66 substitutions per million years is assumed throughout pre-primate evolution (Britten, 1986), from sequence analysis of the serine proteinase domains of HGF/SF, HGFl/MSP, plasminogen, and apolipoprotein (a), it would appear that the divergence of the HGF/SF and HGFl/MSP genes occurred more than 500 million years ago and that the divergence of the plasminogen and apolipoprotein (a) genes occurred at least 23 million years ago. A previous estimate of the divergence of the plasminogen and apolipoprotein (a) genes based on sequences at the 3′ untranslated end of their cDNAs led to the conclusion that the 2 genes had diverged approximately 40 million years ago (McLean et al., 1987). The present study suggests that the loss of proteinase activity of HGF/SF and HGFl/MSP occurred before the divergence of their respective genes, but does not provide an evolutionary explanation for the fact that the N domain of plasminogen is cleaved during activation, whereas the corresponding domains of HGFl/MSP and HGF/SF are not. The

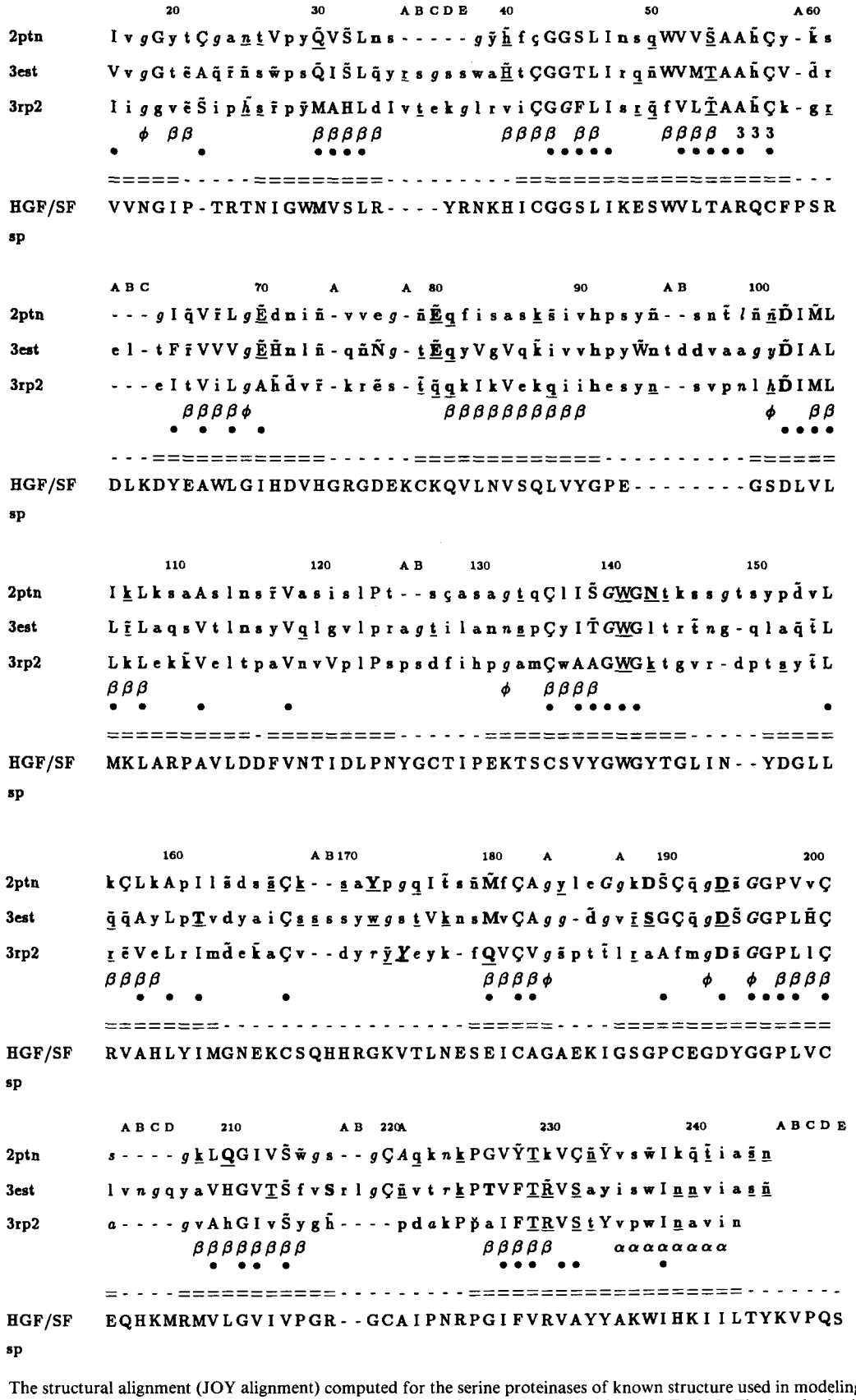

Fig. 7. The structural alignment (JOY alignment) computed for the serine proteinases of known structure used in modeling the serine proteinase domain of HGF/SF and the sequence alignment of the latter. PDB codes as in Table 2. The standard 1-letter code is used with the same additions as those presented in Figure 3. The SCRs (===) and SVRs (———) are marked. 3 denotes 3_{10}-helical main chain. Numbering at the top corresponds to that of 2PTN.

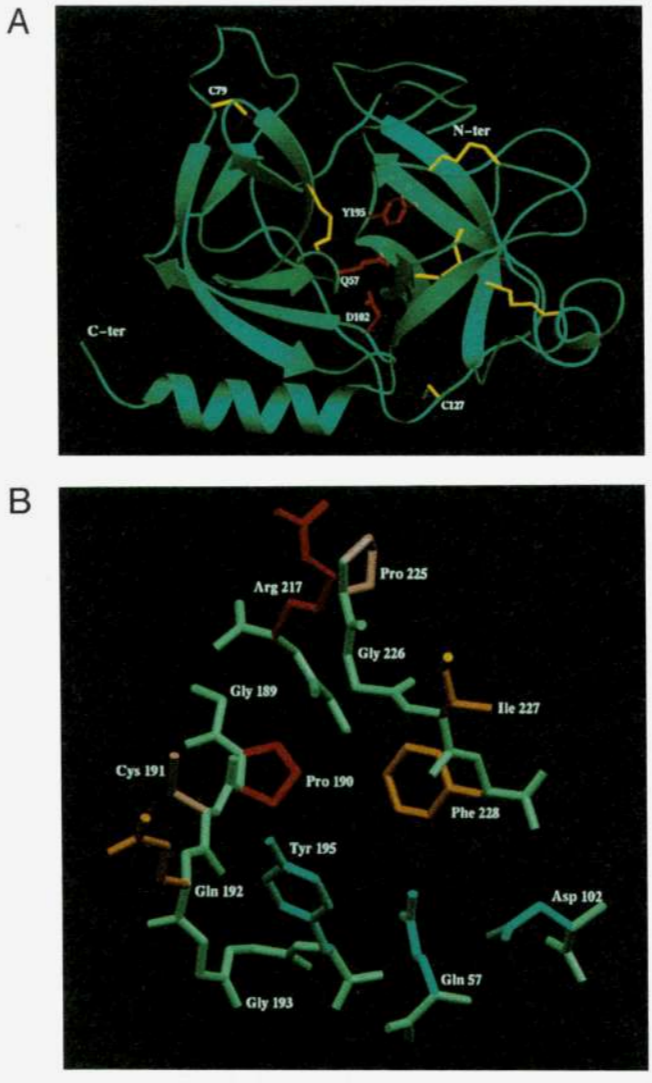

Fig. 8. A: The overall fold of the energy-minimized modeled serine proteinase domain of HGF/SF. The 4 disulfide bonds and the 2 Cys residues, not involved in intradomain disulfide bonds, are shown in yellow. The residues forming the catalytic triad are shown in red, where His_{57} and Ser_{195} have been replaced by Gln and Tyr residues, respectively. **B:** Close-up of the residues forming the S1 specificity pocket of HGF/SF serine proteinase domain. The positions that have experienced drastic replacements, when compared with trypsin, are highlighted in red. The positions of conservative replacements are shown in pink. Invariant residues are shown in green and those of the catalytic triad in blue.

lack of cleavage of the N domain during activation must have been a critical step during the evolution of HGF/SF, because HGF/SF lacking the N domain or its hairpin loop does not bind to the *Met* receptor (Matsumoto et al., 1991; Okigaki et al., 1992). It is conceivable that the N domain of HGFl/MSP is equally critical for activity, but this remains to be established.

To understand the contribution of the different domains of plasminogen-related growth factors to receptor binding and biological activity, we have built models of the critical regions of the N (PAP/PAP homology) domain, based on disulfide bond constraints, and of the 4 kringle domains and the serine proteinase domain of HGF/SF, based on the crystallographic structures of the corresponding homologues.

In the absence of known 3D structures homologous to the hairpin loop of the N-terminal domain, the modeling was approached first by searching for a fold compatible with the constraints dictated by the known disulfide-bonding connectivity. The hairpin loop of the N domain appears to be an independent unit, although the procedure searched through substructures of larger folding units. The number of hits corresponding to a common fold (as was the case here, where 4 of the 11 hits shared the same fold) may reflect its stability even in isolation.

Mutagenesis experiments have shown that HGF/SF mutants lacking most of the N-terminal domain (Okigaki et al., 1992) and lacking the 7 amino acids positioned between the 2 inner Cys residues of the hairpin loop (Mastumoto et al., 1991) are totally inactive. Interestingly, in the structural model corresponding to a helix–extended strand–helix motif proposed here for the N-terminal domain of HGF/SF, an extended strand is formed by these 7 residues; therefore, the proposed model is in good agreement with the experimental data obtained for the deletion mutant. The model of the hairpin loop of HGF/SF indicates an amphipathic distribution of charged and hydrophobic residues on opposite faces of the hairpin. These clusters could play a role in receptor or interdomain binding.

The sequences of all 4 kringles of HGF/SF are similar, particularly to 1PK4, but also to 1TPK. Therefore, most of the SCRs and the SVRs were modeled on 1PK4; this is reflected in the small RMS differences obtained after multiple superposition of the models onto the crystal structures of the homologues. Furthermore, the modeled kringle domains maintain all the general features common to the kringle domains solved by X-ray analysis (such as the triplet of disulfide bonds that determines the very tight fold).

There are 2 forms of HGF/SF kringle 1: one is 79 amino acids in length, and the other (k1Δ) has a deletion of 5 residues and results from alternative splicing of the HGF/SF primary transcript (Seki et al., 1990; Rubin et al., 1991). We have modeled these 2 variants and have shown that the deletion of the 5 amino acids maps onto loop 2 of kringle 1. Because both HGF/SF kringle 1 variants are equally active, it is unlikely that loop 2 of this kringle is involved in HGF/SF binding to its receptor or in HGF/SF general folding.

Because several kringle domains of serine proteinases involved in the removal of blood clots have lysine-binding activity, we have looked for lysine-binding pockets within the modeled kringle domains of HGF/SF. The X-ray structures of the lysine-binding pockets of both 1PK4 and 1TPK indicate the existence of a lysine-binding pocket in the form of a V-shaped trough made by the indole rings of 2 Trp side chains (Trp_{62} and Trp_{72} in 1PK4 and 1TPK). On these grounds, only HGF/SF kringles 2 and 4 are plausible candidates for harboring a lysine-binding pocket.

The 2 other significant features of the lysine-binding pocket in 1PK4 and 1TPK are a cationic cluster (Arg_{32}, Lys_{35}, and Arg_{71}) and an anionic one (Asp_{55} and Asp_{57}) located at opposite ends of the trough. Because in HGF/SF kringle 4 the cationic group no longer exists (it has been replaced by Leu_{32}, His_{35}, and Pro_{71}), we focused our attention on HGF/SF kringle 2 and simulated the docking of ACA (a lysine analogue) into the putative lysine-binding pocket. The ligand could be fitted after minor displacement of Gln_{57} (which replaces Asp_{57} in 1PK4) to avoid steric clashes and a rotation of Phe_{35} (which replaces Lys_{35} in 1PK4) into a buried conformation (in 1PK4 Lys_{35} is exposed

on the surface). The hydrogen bond interaction that exists in the 1PK4–ACA complex between the negatively charged moiety of ACA and Lys_{35} can be maintained in the model with the -OH group of Tyr_{64} (which in 1PK4 is Phe_{64}), although the ionic interaction is obviously lost.

In kringle 4 of HGF/SF, the cationic cluster at one end of the binding trough is lost. This cluster is important for the specificity of ligand binding rather than binding itself (De Serrano & Castellino, 1992a, 1992b). It is conceivable, therefore, that kringle 4 of HGF/SF may accommodate ligands other than lysine into its putative binding pocket.

The serine proteinase domain of HGF/SF was modeled on the 3 closest mammalian serine proteinase structures available in the Brookhaven Protein Data Bank (in terms of sequence homology to HGF/SF). 2PTN and 3EST structures contributed equally, and to a greater extent to the modeled structure than 3RP2, reflecting the fact that the HGF/SF serine proteinase domain has a sequence similarity slightly higher with 2PTN and 3EST than with 3RP2. HGF/SF has no proteolytic activity due to replacement of 2 residues of the active site (His by Gln and Ser by Tyr). Mutant HGF/SF forms have been produced in which the residues of the active site have been reintroduced (Matsumoto et al., 1991; Lokker et al., 1992), but it is unclear whether this is sufficient to restore proteinase activity to the HGF/SF domain. We have focused our attention in HGF/SF to the region corresponding to the S1 specificity pocket of the serine proteinase domain on the grounds that it might be involved in receptor binding or interdomain interaction. The modeled S1 pocket of the HGF/SF serine proteinase domain differs mainly from that of trypsin in that the 2 residues located at the bottom of the pocket (Asp_{189} and Ser_{190}), which are critical for arginine binding, have been substituted by Gly and Pro. Therefore, proper accommodation of arginine at the S1 pocket of the HGF/SF serine proteinase domain does not seem very likely, although the possibility of fitting some other ligand/inhibitor remains to be tested.

Of the 2 Cys residues of the HGF/SF serine proteinase domain that are not involved in disulfides with other Cys residues within the domain, Cys_{127} is involved in the disulfide bridge that links the A and B chains of HGF/SF, whereas no role has been assigned for Cys_{97}.

The *Met* protein, the receptor of HGF/SF (Bottaro et al., 1991; Naldini et al., 1991), is a heterodimer with the 2 subunits disulfide-linked. The α-subunit (45 kDa) is extracellular and the β-subunit (145 kDa) is divided in 3 domains: an extracellular portion, a membrane-spanning region, and an intracellular tyrosine kinase domain. Upon binding of HGF/SF, the receptor undergoes autophosphorylation of the β-subunit. It is here assumed that activation of the receptor occurs through dimerization, which is a consequence of growth factor binding.

Several variant and derivative forms of HGF/SF have been obtained and characterized in terms of their binding to the *Met* protein and their biological activity (Chan et al., 1991; Hartmann et al., 1992; Lokker et al., 1992; Lokker & Godowski, 1993). Thus, a deletion derivative of HGF/SF comprising the N-terminal hairpin loop and kringle 1 (NK1) has been recently shown to enclose a primary determinant of binding to the *Met* protein (Lokker & Godowski, 1993). However, NK1 is inefficient at promoting autophosphorylation of the *Met* receptor and fails to exhibit mitogenic properties even at very high concentrations and, in fact, can act as a potent antagonist in those assays (Lokker & Godowski, 1993). A deletion derivative of HGF/SF bearing the N domain together with the kringle domains 1 and 2 (NK2) binds to the *Met* receptor with an affinity 5-fold lower than that of the intact HGF/SF (Lokker et al., 1992) but 2-fold higher than that of NK1 (Lokker & Godowski, 1993). This NK2 variant does not have mitogenic properties (Lokker et al., 1992) and it has some scattering activity (Hartmann et al., 1992), but the data with respect to its ability to induce tyrosine phosphorylation are disputed (Hartmann et al., 1992; Lokker et al., 1992). Addition of kringles 3 and 4 does not increase the affinity for the receptor. The serine proteinase domain expressed alone neither binds to the *Met* receptor nor has biological activity (Hartmann et al., 1992). On the other hand, substitution mutants in the serine proteinase domain of HGF/SF have been obtained that have a binding capacity similar to that of the unmodified protein but no detectable biological activity (Lokker et al., 1992). For full activation of the downstream signal cascade involved in both the motility and mitogenic responses to HGF/SF binding, the unmodified, full-length HGF/SF protein is required (Hartmann et al., 1992).

For the receptor to dimerize, the ligand should either possess 2 binding sites or be a dimer in itself. Because the HGF/SF derivative lacking the first 187 N-terminal amino acids (N domain plus kringle 1) does not have any binding capacity, it seems unlikely that 2 receptor-binding sites exist within the HGF/SF molecule. There are certain features of HGF/SF, such as the free Cys residues in kringles 2 and 3 as well the one in the serine proteinase domains, that make the possibility of HGF/SF forming homodimers rather appealing. However, there is strong experimental evidence that HGF/SF does not exist as a covalently linked homodimer in solution, ruling out the possibility of intermolecular disulfide bonds between Cys_4 of kringle 2, Cys_{44} of kringle 3, and Cys_{110} of the serine proteinase domain.

We suggest instead that the activation of the HGF/SF receptor might be mediated through binding of a noncovalently linked HGF/SF homodimer. This proposed dimer would be maintained in solution through interactions involving kringle 2, kringle 3, and the serine proteinase domain, as outlined in Figure 9. In each of the molecules of this putative HGF/SF homodimer, kringle 2 and kringle 3 are assumed have been brought together through a disulfide bond established between Cys_4 of kringle 2 and Cys_{44} of kringle 3. This is made feasible by the length of the connecting peptide (16 residues long) and its amino acid sequence, which allows flexibility for the kringle domains to move with respect to each other. Kringle 2 of molecule 1 could then accommodate a lysine residue from kringle 3 of molecule 2 (it could be either Lys_{43} or Lys_{45}, the only 2 Lys residues in the sequence of kringle 3) into its putative lysine-binding pocket and, vice versa, kringle 2 of molecule 2 could interact with a Lys residue from kringle 3 of molecule 1. This type of ligand-like binding interaction between kringles has been recently reported in t-PA kringle 2 crystals (Padmanabhan et al., 1994). The putative noncovalent HGF/SF homodimer might also be held through direct interactions between the serine proteinase domains of each molecule mediated by the catalytic cleft and/or the S1 pocket; this could accommodate residues, other than arginine, from the serine proteinase domain of one of the HGF/SF molecules. Although kringle 4 would not be needed directly to sustain the HGF/SF homodimer in this model, it could contribute to the maintenance of the required conformation of the HGF/SF molecule for proper dimerization, perhaps by interacting with kringle 1 through its proposed lysine binding-like pocket. The model

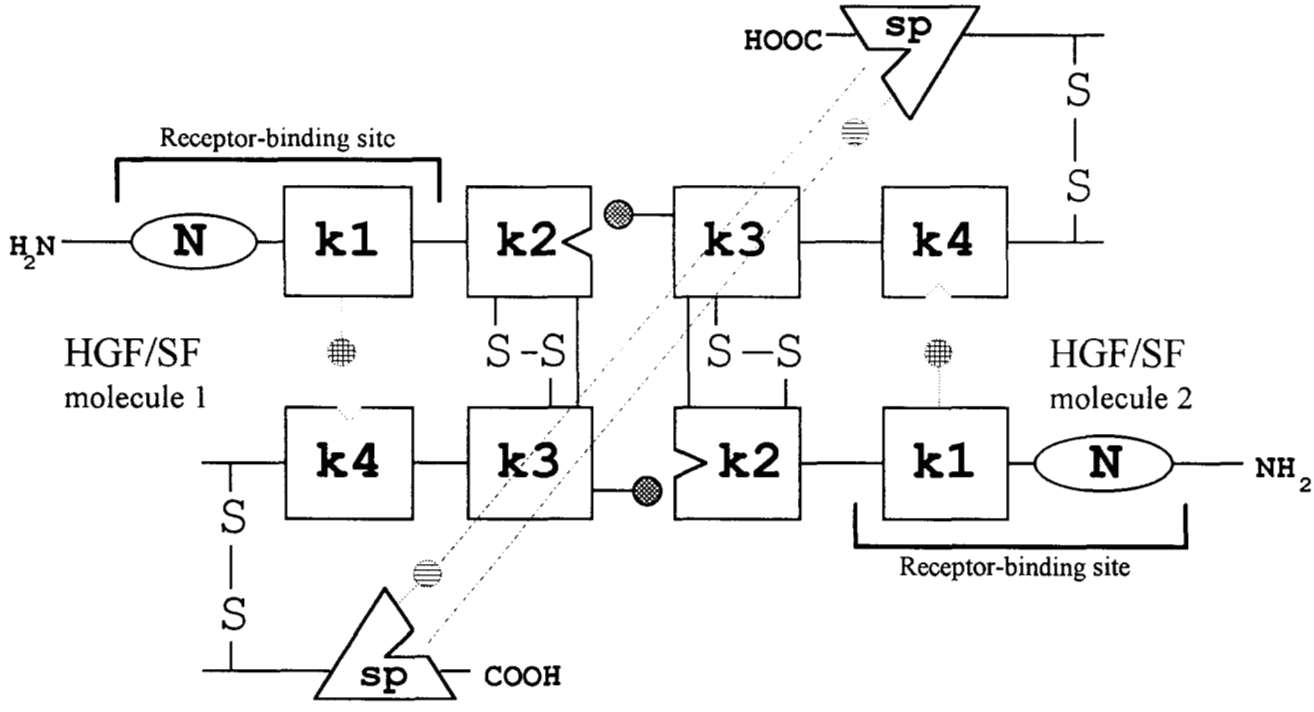

Fig. 9. Schematic drawing of the proposed HGF/SF noncovalent homodimer. A model of the proposed noncovalent HGF/SF homodimer required for dimerization and activation of the receptor *Met* protein. k*n*, kringle domain *n*; sp, serine proteinase domain.

proposed provides a working hypothesis for HGF/SF-induced activation of the *Met* receptor that can now be tested experimentally by site-directed mutagenesis.

Materials and methods

Sequences were extracted from the EMBL and SwissProt databases and aligned using CLUSTAL V (Higgins & Sharp, 1988). The PHYLIP package (version 3.5) was used to calculate phylogenies from the DNA sequences of the B chains (serine proteinase domains). Transition-transversion ratios were left at the default values, and gaps were replaced with the ambiguity code "?" implying a nucleotide might or might not be present. Trees were calculated from cDNA alignments only and the input order of test sequences was jumbled 10 times. All methods used (parsimony, maximum-likelihood, Fitch-Margoliash least squares) produced topologically identical trees.

Comparative modeling of the 3D structures of the kringle domains and serine proteinase domain of HGF/SF has been performed by the procedure encoded in the suite of computer programs COMPOSER (Blundell et al., 1988).

In COMPOSER, the amino acid sequence of an unknown structure is first matched with sequences of one or more homologous 3D structures from the Brookhaven Protein Data Bank (Bernstein et al., 1977). In practice, homologues with a sequence identity of >30% are usually selected. The tertiary structures of the homologues are superposed and the SCRs of the homologous family of proteins are defined by a distance cutoff (2.5 Å) between the topologically equivalent α-carbon positions. Weighted mean positions of topologically equivalent positions in the SCRs constitute a framework for the family of proteins (Sutcliffe et al., 1987a).

The amino acid sequence of the unknown is aligned to the framework, to identify equivalent regions. The SCR of highest local sequence similarity with the unknown is superimposed on the framework to model the corresponding SCR of the unknown. Fragments to model the SVRs are first selected on the basis of the end-to-end distances of the SCRs (Blundell et al., 1988; Topham et al., 1990). The work of Topham et al. (1993) provides a rational procedure for loop selection, based on structural templates of the search segment using amino acid substitution tables (Overington et al., 1990, 1992). The best fragment is "melded" with the SCRs after making small alterations of torsion angles to match distances in the SCRs with the corresponding distances in the fragment picked (F. Eisenmenger, unpubl. results). van der Waals clashes between the SVR and other parts of the protein are also important considerations in loop selection.

The side chains are modeled using a rule-based procedure (Sutcliffe et al., 1987b) that depends on the identity of the amino acid, the secondary structure, and the orientation of equivalent side chains in homologous proteins.

To improve the accuracy of the COMPOSER-built model of the serine proteinase domain of HGF/SF, a weighting factor has been used to derive the framework. This weight is a function of sequence similarity and assumes inverse proportionality between the square of sequence identity and the RMS differences of Cα atoms for the superposed structures used to build the model (Srinivasan & Blundell, 1993).

A complex between the model of HGF/SF kringle 2 and ACA was simulated by superposing the coordinates of the crystal structure of the ACA complex of human plasminogen kringle 4 (Wu et al., 1991) and those of the modeled kringle 2 using the program MNYFIT (Sutcliffe et al., 1987a). For proper accommodation of the ligand, minor adjustments of the side chains were carried out using FRODO (Jones, 1982).

The 3D modeling of the N-terminal hairpin loop of HGF/SF, HGFl/MSP, and plasminogen, which consists of 2 disulfide bonds in a stretch of about 25 residues, was approached by using disulfide bond constraints, because no homologous protein with known 3D structure is available. It has been shown that the 3D modeling of small systems rich in disulfides is feasible starting from the amino acid sequence and the Cys connectivity pattern because the overall fold of such systems is largely dictated by the presence of covalent crosslinks (Sowdhamini et al., 1993). Here we search a large number of known 3D structures to identify segments containing 2 disulfides with a connectivity pattern as in the N-terminal hairpin loop domains (Sowdhamini, 1992; R. Sowdhamini & P. Balaram, unpubl. results). For this purpose, a database of disulfides is set up. The segments identified are potential candidates for the searched fold.

A database of 141 largely nonhomologous proteins of known 3D structure was first considered, but this contained only 202 disulfides. Because it is desirable to have the size of the database as large as possible, the procedure MODIP (Sodwhamini et al., 1989) to enhance the size of the database was used that can choose sites where strainless disulfides can be introduced into proteins of known 3D structure by site-directed mutagenesis. Therefore, the derived disulfide database contains both natural as well as modeled disulfide bonds. The program SSDBASE, which uses MODIP, was employed to identify and organize the database of disulfide bonds. Apart from the 202 native disulfides, 9,738 disulfides could be modeled using MODIP, thus resulting in a considerable increase in the database. The database was searched, using the program SSPATSEA, for 2 disulfides with loop within loop topology. Because the number of residues between the second and third Cys residues varies in HGF/SF, HGFl/MSP, and plasminogen, variation in length was allowed in the search procedure.

Once such a segment was selected, it was used as the initial structure for the COMPOSER-based 3D modeling of the N-terminal hairpin loop of HGF/SF, HGFl/MSP, and plasminogen. The models were energy minimized using the MAXIMIN2 option in SYBYL using the AMBER force field (Weiner et al., 1984; Singh et al., 1986). During energy minimization, the lone pairs and hydrogen atoms were included. For every run of energy minimization, 20 cycles of Simplex method and a further 100 cycles of Powell algorithm were employed. During the initial cycles, the backbone atoms were maintained fixed and the electrostatic energy term was not considered. This process is primarily employed to relieve the short contacts and to correct geometry, particularly at the "anchor" regions of the various loops. Some of the regions with significant deviation from ideal geometry are rectified by energy minimization of that local region. At every stage of energy minimization, the stereochemical quality of the model is checked. In the final stages of energy minimization, when almost all problems related to steric clashes and bad geometry are sorted out, all atoms were allowed to move and the electrostatic energy term was switched on. A distance-dependent dielectric constant with the distance cutoff of 9 Å and a value of 4 were used. Minimization was carried out until all inconsistencies in geometry rectified and all the short contacts were relieved.

All computations were performed on an IRIS 4D workstation, using the COMPOSER module available in SYBYL (version 5.5; Tripos Associates, Inc.). Coordinates used in this study were refined to a resolution of 2.5 Å or better, with the exception of bovine hydrolase and the Mengo virus coat proteins structures, which were solved at resolutions of 2.7 Å and 3.0 Å, respectively.

Acknowledgments

This work has been supported by the Imperial Cancer Research Fund. L.E.D. is recipient of a fellowship from the Human Capital and Mobility program of the European Community. N.S. is supported by a grant from Tripos Associates, Inc.

References

Arai KI, Lee F, Miyajima A, Miyatake S, Arai N, Yokota T. 1990. Cytokines: Coordinators of immune and inflammatory responses. *Annu Rev Biochem 59*:783–836.

Baron M, Norman DG, Campbell ID. 1991. Protein modules. *Trends Biochem Sci 16*:13–17.

Bernardi G. 1989. The isochore organization of the human genome. *Annu Rev Genet 23*:637–661.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Bezerra JA, Han S, Danton MJS, Degen SIF. 1993. Are hepatocyte growth factor-like protein and macrophage stimulating protein the same protein? *Protein Sci 2*:666–668.

Blundell TL, Carney D, Gardner S, Hayes F, Howlin B, Hubbard T, Overington J, Singh DA, Sibanda BL, Sutcliffe M. 1988. 18th Hans Krebs Lecture. Knowledge-based protein modelling and design. *Eur J Biochem 172*:513–520.

Bode W, Turk D, Karshikov A. 1992. The refined 1.9 Ångstrom X-ray crystal structure of D-Phe-Pro-Arg-chloromethylketone inhibited human α-thrombin. Structure analysis, overall structure, electrostatic properties, detailed active site geometry and structure-function relationships. *Protein Sci 1*:426–471.

Bottaro DP, Rubin JS, Faletto DL, Chan AML, Kmiecick T, Vande Woude G, Aaronson SA. 1991. Identification of the hepatocyte growth-factor receptor as the c-Met protooncogene product. *Science 251*:802–804.

Britten RJ. 1986. Rates of DNA sequence evolution differ between taxonomic groups. *Science 231*:1393–1398

Burgess WH, Maciag T. 1989. The heparin-binding (fibroblast) growth-factor family of proteins. *Annu Rev Biochem 58*:575–606.

Burley SK, David PR, Taylor AN, Lipscomb WN. 1990. Molecular structure of leucine aminopeptidase at 2.7 Ångstrom resolution. *Proc Natl Acad Sci USA 87*:6878–6882.

Chan A, Rubin JS, Bottaro DP, Hirschfield DW, Chedid M, Aaronson SA. 1991. Identification of a competitive HGF antagonist encoded by an alternative transcript. *Science 254*:1382–1384.

Daopin S, Cohen G, Davies DR. 1992. Structural similarity between transforming growth-factor-beta-2 and nerve growth-factor response. *Science 258*:1160–1162.

Deisenhofer J, Michel H. 1989. The photosynthetic reaction center from the purple bacterium *Rhodopseuodomonas viridis*. *Science 245*:1463–1473.

Derynck R. 1988. Transforming growth factor alpha. *Cell 54*:593–595.

De Serrano VS, Castellino FJ. 1992a. Role of tryptophan 74 of the recombinant kringle-2 domain of tissue-type plasminogen activator in its ω-amino acid binding properties. *Biochemistry 31*:3326–3335.

De Serrano VS, Castellino FJ. 1992b. The cationic locus of the recombinant kringle-2 domain of tissue-type plasminogen activator that stabilizes its interaction with ω-amino acids. *Biochemistry 31*:11698–11706.

De Vos AM, Ultsch MH, Kelley RF, Padmanabhan K, Tulinsky A, Westbrook ML, Kossiakoff AA. 1992. Crystal structure of the kringle 2 domain of tissue plasminogen activator at 2.4 Ångstrom resolution. *Biochemistry 31*:270–279.

Doolittle WF. 1987. What introns have to tell us. Hierarchy in genome evolution. *Cold Spring Harbor Symp Quant Biol LII*:907–913.

Dorit RL, Schoenbach L, Gilbert W. 1990. How big is the universe of exons? *Science 250*:1377–1382.

Ebendal T. 1992. Function and evolution in the NGF family and its receptors. *J Neurosci Res 33*:461–470.

Fita I, Rossmann MG. 1985. The NADPH binding-site on beef liver catalase. *Proc Natl Acad Sci USA 82*:1604–1608.

Gherardi E, Gray J, Stoker M, Perryman M, Furlong R. 1989. Purification of scatter factor, a fibroblast derived basic protein that modulates epithelial interactions and movement. *Proc Natl Acad Sci USA 86*:5844–5848.

Gill GN, Bertics PJ, Santon JB. 1987. Epidermal growth-factor and its receptor. *Mol Cell Endocrinol 15*:169–186.

Golding B, Felsenstein J. 1990. A maximum likelihood approach to the detection of selection from a phylogeny. *J Mol Evol 31*:511–523.

Han S, Stuart LA, Degen SJF. 1991. Characterization of the DNF15S2 locus on human chromosome 3. Identification of a gene coding for 4 kringle domains with homology to hepatocyte growth factor. *Biochemistry 30*:9768–9780.

Hartmann G, Naldini L, Weidner KM, Sachs M, Vigna E, Comoglio PM, Birchmeier W. 1992. A functional domain in the heavy chain of scatter factor hepatocyte growth-factor binds the c-Met receptor and induces cell dissociation but not mitogenesis. *Proc Natl Acad Sci USA 89*:11574–11578.

Heldin CH, Ostman A, Westermark B. 1993. Structure of platelet-derived growth factor. Implications for functional properties. *Growth Factors 8*:245–252.

Higgins DG, Sharp PM. 1988. Clustal, a package for performing multiple sequence alignment on a microcomputer. *Gene 73*:237–244.

Huff JL, Jelinek MA, Borgman CA, Lansing TJ, Parsons JT. 1993. The protooncogene c-Sea encodes a transmembrane protein-tyrosine kinase related to the Met hepatocyte growth-factor scatter factor receptor. *Proc Natl Acad Sci USA 90*:6140–6144.

Humbel RE. 1990. Insulin-like growth factor-I and factor-II. *Eur J Biochem 190*:445–462.

Jones TA. 1982. FRODO: A graphics fitting program for macromolecules. In: Sayre D, ed. *Computational crystallography*. Oxford, UK: Clarendon Press. pp 303–317.

Krishnaswamy S, Rossmann MG. 1990. Structural refinement and analysis of Mengo virus. *J Mol Biol 211*:803–844.

Lokker NA, Godowski PJ. 1993. Generation and characterization of a competitive antagonist of human hepatocyte growth-factor, HGF/NK1. *J Biol Chem 268*:17145–17150.

Lokker NA, Mark MR, Luis EA, Bennett GL, Robbins KA, Baker JB, Godowski PJ. 1992. Structure-function analysis of hepatocyte growth-factor. Identification of variants that lack mitogenic activity yet retain high-affinity receptor-binding. *EMBO J 11*:2503–2510.

Mars WM, Zarnegar R, Michalopoulos GK. 1993. Activation of hepatocyte grwth-factor by the plasminogen activators uPA and tPA. *Am J Pathol 143*:949–958.

Massagué J. 1990. The transforming growth-factor beta family. *Annu Rev Cell Biol 6*:597–641.

Matsumoto K, Takehara T, Inoue H, Hagiya M, Shimizu S, Nakamura T. 1991. Deletion of kringle domains or the N-terminal hairpin structure in hepatocyte growth factor results in marked decreases in related biological activities. *Biochem Biophys Res Commun 181*:691–699.

McLean JW, Tomlinson JE, Kuang J, Eaton DL, Chen EY, Fless GM, Scanu AM, Lawn RM. 1987. cDNA sequence of human apolipoprotein (a) is homologous to plasminogen. *Nature 300*:132–137.

Metcalf D. 1989. The molecular control of cell division, differentiation commitment and maturation in hematopoietic cells. *Nature 339*:27–30.

Meyer E, Cole G, Radhakrishnan R, Epp O. 1988. Structure of native porcine pancreatic elastase at 1.65 Ångstrom resolution. *Acta Crystallogr B 44*:26–38.

Miyazawa K, Shimomura T, Kitamura A, Kondo J, Morimoto Y, Kitamura N. 1993. Molecular cloning and sequence analysis of the cDNA for a human serine protease responsible for activation of hepatocyte growth-factor. Structural similarity of the protease precursor to blood coagulation factor XII. *J Biol Chem 268*:10024–10028.

Miyazawa K, Tsubouchi H, Naka D, Takahashi K, Okigaki M, Arakaki N, Nakayama H, Hirono S, Sakiyama O, Takahashi K, Godha E, Daikuhara Y, Kitamura N. 1989. Molecular cloning and sequence analysis of cDNA for human hepatocyte growth-factor. *Biochem Biophys Res Commun 163*:967–973.

Mulichak AM, Tulinsky A, Ravichandran KG. 1991. Crystal and molecular structure of human plasminogen kringle 4 refined at 1.9 Ångstrom resolution. *Biochemistry 30*:10576–10588.

Murray-Rust J, McDonald NQ, Blundell TL, Hosang M, Oefner C, Winkler F, Bradshaw RA. 1993. Topological similarities in TGG-beta-2, PDGF-BB and NGF define a superfamily of polypeptide growth-factors. *Structure 1*:153–159.

Murzin AG, Chothia C. 1992. Protein architecture: New superfamilies. *Curr Opin Struct Biol 2*:895–903.

Nakamura T, Nishizawa T, Hagiya M, Seki T, Shimonishi M, Sugimura A, Tashiro K, Shimizu S. 1989. Molecular cloning and expression of human hepatocyte growth factor. *Nature 342*:440–443.

Naldini L, Tamagnone L, Vigna E, Sachs M, Hartmann G, Birchmeier W, Daikuhara Y, Tsubouchi H, Blasi F, Comoglio PM. 1992. Extracellular proteolytic cleavage by urokinase is required for activation of hepatocyte growth-factor scatter factor. *EMBO J 11*:4825–4833.

Naldini L, Weidner KM, Vigna E, Gaudino G, Bardelli A, Ponzetto C, Narsimhan RP, Hartmann G, Zarnegar R, Michalopoulos GK. 1991. Scatter factor and hepatocyte growth-factor are indistinguishable ligands for the Met receptor. *EMBO J 10*:2867–2878.

Okigaki M, Komada M, Uehara Y, Miyazawa K, Kitamura N. 1992. Functional characterization of human hepatocyte growth-factor mutants obtained by deletion of structural domains. *Biochemistry 31*:9555–9561.

Overington JP, Donnelly D, Johnson MS, Sali A, Blundell TL. 1992. Environment specific amino acid substitution tables, tertiary templates, and prediction of protein folds. *Protein Sci 1*:216–226.

Overington JP, Johnson MS, Sali A, Blundell TL. 1990. Tertiary structural constraints on protein evolutionary diversity templates, key residues and structure prediction. *Proc R Soc (Lond) 241*:132–145.

Padmanabhan K, Wu TP, Ravichandran KG, Tulinsky A. 1994. Kringle-kringle interactions in multimer kringle structures. *Protein Sci 3*:898–910.

Reid KSC, Lindley PF, Thornton JM. 1985. Sulfur-aromatic interactions in proteins. *FEBS Lett 190*:209–217.

Remington SJ, Woodbury RG, Reynolds RA, Matthews BW, Neurath H. 1988. The structure of rat mast-cell protease-II at 1.9 Ångstrom resolution. *Biochemistry 27*:8097–8105.

Ronsin C, Muscatelli F, Mattei MG, Breathnach R. 1993. A novel putative receptor protein tyrosine kinase of the Met family. *Oncogene 8*:1195–1202.

Rubin JS, Chan AML, Bottaro DP, Burgess W, Taylor WG, Cech AC, Hirschfield DW, Wong J, Miki T, Finch PW, Aaronson SA. 1991. A broad-spectrum human lung fibroblast derived mitogen is a variant of hepatocyte growth-factor. *Proc Natl Acad Sci USA 88*:415–419.

Šali A, Blundell TL. 1990. Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *J Mol Biol 212*:403–428.

Seki T, Hagiya M, Shimonishi M, Nakamura T, Shimizu S. 1991. Organization of the human hepatocyte growth-factor encoding gene. *Gene 102*:213–219.

Seshadri TP, Tulinsky A, Skrzypczak-Jankun E, Park CM. 1991. Structure of bovine prothrombin fragment-1 refined at 2.25 Ångstrom resolution. *J Mol Biol 220*:481–494.

Singh UC, Brown FK, Bash PA, Kollman PA. 1987. An approach to the application of free-energy perturbation methods using molecular dynamics. Application to the transformations of $CH_3OH \rightarrow CH_3CH_3$, $H_3O^+ \rightarrow NH_4^+$, glycine → alanine, and alanine → phenylalanine in aqueous solution and to $H_3O^+(H_2O)_3 \rightarrow NH_4^+(H_2O)_3$ in the gas phase. *J Am Chem Soc 109*:1607–1614.

Skeel A, Yoshimura T, Showalter SD, Tanaka S, Appella E, Leonard EJ. 1991. Macrophage stimulating protein. Purification, partial amino acid sequence and cellular activity. *J Exp Med 173*:1227–1234.

Sowdhamini R. 1992. Motifs in proteins: Disulfide constraints and their applications to protein engineering and peptide modeling. [thesis]. Bangalore, India: Indian Institute of Science.

Sowdhamini R, Ramakrishnan C, Balaram P. 1993. Modeling multiple disulfide loop containing polypeptides by random conformation generation. The test case of alpha-conotoxin GI and endothelin-I. *Protein Eng 6*:873–882.

Sowdhamini R, Srinivasan N, Shoichet B, Santi DV, Ramakrishnan C, Balaam P. 1989. Stereochemical modelling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis. *Protein Eng 3*:95–103.

Srinivasan N, Blundell TL. 1993. An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. *Protein Eng 6*:501–512.

Stoker M, Gherardi E, Perryman M, Gray J. 1987. Scatter factor is a fibroblast-derived modulator of epithelial cell mobility. *Nature 327*:239–242.

Sutcliffe MJ, Haneef I, Carney D, Blundell TL. 1987a. Knowledge-based modelling of homologous proteins. 1. 3-Dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng 1*:377–384.

Sutcliffe MJ, Hayes FRF, Blundell TL. 1987b. Knowledge-based modelling of homologous proteins. 2. Rules for the conformations of substituted side chains. *Protein Eng 1*:385-392.

Topham CM, McLeod A, Eisenmenger F, Overington JP, Johnson MS, Blundell TL. 1993. Fragment ranking in modelling of protein structure. Conformationally constrained environmental amino acid substitution tables. *J Mol Biol 229*:194-220.

Topham CM, Thomas P, Overington JP, Johnson MS, Eisenmenger F, Blundell TL. 1990. An assessment of COMPOSER, a rule-based approach to modeling protein structure. *Biochem Soc Symp 57*:1-9.

Ullrich A, Schlessinger J. 1990. Signal transduction by receptors with tyrosine kinase activity. *Cell 21*:203-212.

Walter J, Steigemann W, Singh TP, Bartunik H, Bode W, Huber R. 1982. On the disordered activation domain in trypsinogen. Chemical labeling and low-temperature crystallography. *Acta Crystallogr B 38*:1462-1472.

Weidner KM, Arakaki N, Hartmann G, Vandekerckhove J, Weingart S, Rieder H, Fonatsch C, Tsubouchi H, Hishida T, Daikuhara Y, Birchmeier W. 1991. Evidence for the identity of human scatter factor and human hepatocyte growth factor. *Proc Natl Acad Sci USA 88*:7001-7005.

Weidner KM, Behrens J, Vandekerckhove J, Birchmeier W. 1990. Scatter factor. Molecular characteristics and effect on the invasiveness of epithelial cells. *J Cell Biol 111*:2097-2108.

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Algona G, Profeta S, Weiner PK. 1984. A new force-field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc 106*:765-784.

Welch HM, Darby JK, Pilz AJ, Ko CM, Carritt B. 1989. Transposition, amplification and divergence in the origin of the DNF15 loci, a polymorphic repetitive sequence family on chromosome-1 and chromosome-3. *Genomics 5*:423-430.

Wiman B. 1973. Primary structure pf peptides released during activation of human plasminogen by urokinase. *Eur J Biochem 39*:1-9.

Wu TP, Padmanabhan K, Tulinsky A, Mulichak AM. 1991. The refined structure of the ϵ-aminocaproic acid complex of human plasminogen kringle 4. *Biochemistry 30*:10589-10594.

Yoshimura T, Yuhki N, Wang MH, Skeel A, Leonard EJ. 1993. Cloning, sequencing and expression of human macrophage stimulating protein (MSP, MST1) confirms MSP as a member of the family of kringle proteins and locates the MSP gene on chromosome-3. *J Biol Chem 268*: 15461-15468.

Zhu ZY, Sali A, Blundell TL. 1992. A variable gap penalty function and feature weights for protein 3-D structure comparisons. *Protein Eng 5*:43-51.