The evolution of vertebrate blood coagulation as viewed from a comparison of puffer fish and sea squirt genomes

Yong Jiang and Russell F. Doolittle*

Center for Molecular Genetics, University of California at San Diego, La Jolla, CA 92039-0634

Contributed by Russell F. Doolittle, May 1, 2003

The blood coagulation scheme for the puffer fish, Fugu rubripes, has been reconstructed on the basis of orthologs of genes for mammalian blood clotting factors being present in its genome. As expected, clotting follows the same fundamental pattern as has been observed in other vertebrates, even though genes for some clotting factors found in mammals are absent and some others are present in more than one gene copy. All told, 26 different proteins involved in clotting or fibrinolysis were searched against the puffer fish genome. Of these, orthologs were found for 21. Genes for the "contact system" factors (factor XI, factor XII, and prekallikrein) could not be identified. On the other hand, two genes were found for factor IX and four for factor VII. It was evident that not all four factor VII genes are functional, essential active-site residues having been replaced in two of them. A search of the genome of a urochordate, the sea squirt, Ciona intestinalis, did not turn up any genuine orthologs for these 26 factors, although paralogs and/or constituent domains were evident for virtually all of them.

B

lood clotting follows the same fundamental pattern in all vertebrates, from the early diverging jawless fishes to mammals (1). In all cases the principal event is the thrombin-catalyzed conversion of a soluble plasma protein, fibrinogen, into an insoluble polymeric fibrin clot. Thrombin is a serine protease, itself the product of a series of proteolytic events. It is well established that all groups of fish (cyclostomes, clasmobranches, and teleosts) generate thrombin by pathways involving vitamin K-dependent factors, exhibit factor XIII-dependent fibrin cross-linking, and manifest a fibrinolysis that is inhibited by the same agents as inhibit fibrinolysis in mammals (1–3). In contrast, thrombin-generated fibrin clotting has not been reported in nonvertebrate chordates or other invertebrate animals.

Because such a convoluted pathway could not have evolved in one fell swoop, it was long ago realized that a series of gene duplications must lie at the heart of the complex set of interactions observed in mammalian clotting. In this regard, past sequence comparisons of serine proteases have led to the suggestion that certain of the clotting factors (particularly those constituting the "contact system" involving factors XI and XII, and prekallikrein) must have made their appearance more recently in evolution than some of the other clotting factors and would likely be absent in lower vertebrates (4).

The recent publication (5) of the genome sequence for the puffer fish, Fugu rubripes, makes it possible to test that prediction. Additionally, the availability (6) of the complete genome sequence for a urochordate, the sea squirt, Ciona intestinalis, allows a direct comparison of two early diverging chordates. The results confirm that the main lines of the vertebrate clotting pathway were evolved in the interval between the last common ancestor of these two creatures, a period thought to be significantly less than a hundred million years.

In the present study, 26 proteins involved in mammalian blood clotting (Table 1) were examined to see whether they have counterparts in the puffer fish and/or the sea squirt. The processes of fibrin formation and destruction are inextricably linked, and the proteins selected include both lytic factors and inhibitors. Because many paralogs could be involved (the result of recent gene duplications), stringent criteria were set for deciding whether or not a gene for a given coagulation factor was present. In the end, 21 orthologs of the coagulation factor genes were found in the puffer fish genome, but not one authentic ortholog was identified in the sea squirt genome.

On the other hand, the wherewithal for the evolutionary assembly of all of the coagulation factor genes seems to be present in the sea squirt genome in the form of various and sundry domains that in ancestral lines were duplicated and rearranged into the genes for the present-day vertebrate clotting factors. Many of these accessory domains occur at the amino terminii of zymogens for serine proteases. Paramount among them is the "GLA domain," an entity that contains multiple γ-carboxy-glutamic acid residues whose synthesis depends on vitamin K (7). The blood coagulation proteins that contain GLA domains include prothrombin, factors VII, IX, and X, protein C, and the nonprotease protein S. Other peripheral domains found in association with clotting proteases include "kringle" (8), fibronectin "finger" domains (FN-1 and FN-2; refs. 9 and 10), fibronectin domain III (FN-3; ref. 11), epidermal growth factor (EGF) domains (9, 12), and "apple domains," four copies of which are found at the amino terminii of mammalian factor XI and prekallikrein (13). Apple domains are actually part of a larger group now referred to as PAN domains (14). Additionally, there are sundry other domains associated with clotting proteins, including the F5/8-A and -C domains found in factors V and VIII, two proteins that themselves are evolved from ceruloplasmin (15), although the latter protein is composed exclusively of three A domains. The F5/8-C domain is also known as "discoidin."

Many of these accessory domains are active in localizing the clotting process. GLA domains, for example, serve to bind the vitamin K-dependent factors to platelets or thrombocytes, and other domains serve to bind the factors to each other. In passing, it should be noted that platelets are not found in nonmammals, their role being enacted by a class of white cells designated "thrombocytes."

Methods

Publicly accessible sequence data were downloaded from the Fugu and Ciona web sites: fugu.hgmp.mrc.ac.uk and www.jgi.doe.gov/ciona.

Puffer Fish. The puffer fish DNA data are available in the form of 12,381 scaffolds ranging in size from 657 to 2 Kb. These
sequences amount to 332.5 Mb containing ~30,000 potential genes and make up >95% of the puffer fish genome (6). Candidate genes were initially identified by searching the sequence of each human coagulation protein against the puffer fish scaffolds with the TBLASTN program (16).

In each case, a prioritized list of candidate scaffolds was examined. The approximate matching regions were extracted from scaffolds, and the program GENESCAN (17) was used in an effort to link up exons. In general, GENESCAN was ~75% effective, missing about a quarter of all exons. The roughly translated sequence was aligned with the appropriate human protein, after which missing regions were identified with the aid of the program INSPECT (18). Standard BLAST (16) was then used to back-search the full amino sequence against GenBank. If the candidate was indeed the ortholog of the factor under study, the same factor should be returned (T = True). If that was not the case (F = False), the process was repeated on the next candidate on the list, and so on, until a reasonable conclusion about presence (T) or absence (F) could be reached. The method is not foolproof. Occasionally the best match found in the puffer fish genome gave rise to a much better match upon back-searching, indicating that the candidate was not an ortholog. In some other cases, the matches were reciprocal, but the resemblance was not particularly strong. In this regard, most of the genuine orthologs were >40% identical with their human counterparts (Table 1). Valid orthologs were also expected to have the same arrangement of modular components. Intron locations were also compared. Additionally, for those factors that belong to well characterized families, phylogenetic trees were constructed (19). Pairwise alignments of 21 putative puffer fish proteins with human counterparts and with intron positions marked are published as supporting information on the PNAS web site, www.pnas.org.

Sea Squirt. A similar strategy for the same 26 factors was used for searching the sea squirt genome. The 160-Mb genome encom-
passes 16,000 genes assembled into 2,501 scaffolds (7). The downloaded sequences were subjected to the regimen described above with tblastn (16), genescan (17), inspect (18), and finally blast to back-search candidates against genbank. The same criteria for orthologs were used as applied to puffer fish: namely, (i) the highest resemblance to any protein in genbank with the back-search had to be the coagulation factor, and (ii) the modular arrangement of domains had to be the same as occurs in the mammalian factor.

Results

Puffer fish. Vitamin K-dependent factors. Orthologous genes were identified for each of the five vitamin K-dependent serine proteases (prothrombin, factors VII, IX, and X, and protein C). Remarkably, two genes were found for factor IX, both of which may be functional. The two proteins are 51% identical with each other: a stronger resemblance than is observed with either to mammalian factor IX, indicating that the gene duplication leading to the two homologs occurred after fish diverged from the lineage leading to tetrapods.

Four genes (not all intact) were found for factor VII, all on a single scaffold together with a gene for factor X (Fig. 1). Of the four, the homolog labeled f7B most likely represents the genuine factor, its putative protein product being 46% identical with mammalian factor VII. Homologs f7A and f7C are 42% and 43% identical with the mammalian factor VII, respectively; in homolog f7A, the expected active-site serine codon has been changed to that of aspartic acid. The gene denoted f7F7 in Fig. 1 is in an obvious state of evolutionary decay, the sequence in the region of the expected active site being greatly fragmented. The single factor X gene is 41% identical to human factor X. That the puffer fish factor X and factor VII genes are adjacent to each other (Fig. 1) may be significant in that, among all of the major human coagulation factors, only these two are found at the same locus on the same chromosome (20).

An additional GLA-containing homolog with two EGF domains was found on scaffold 6546, the sequence of which doesn’t fall within any of the four subfamilies with this arrangement of domains (factors VII, IX, or X, or protein C). The gene contains a canonical signal sequence, and its activation site and active site are both intact. All indications are that this could be a fully active gene for the precursor of a serine protease. In contrast, a GLA-containing gene on scaffold 468 most closely resembles factor X, but it lacks the critical active-site constellation of residues needed for catalysis, as well as the key basic-residue cleavage site needed for activation.

A phylogenetic tree based on the serine protease portions of the vitamin K-dependent proteases from human, bovine, and puffer fish clustered all except one of the puffer fish factors with their mammalian counterparts, the one exception being the unassigned factor on scaffold 6546 (Fig. 2). The tree indicates that most of the extra genes for the various factors are the results of duplications that have occurred after fish and tetrapods diverged.

Protein S. The GLA-containing steroid-binding protein known as protein S (21) was present, and like its human counterpart, it is situated near a similar homolog, which in humans is denoted as a growth arrest-specific gene (22). The single factor X gene is 41% identical to human factor X. That the puffer fish factor X and factor VII genes are adjacent to each other (Fig. 1) may be significant in that, among all of the major human coagulation factors, only these two are found at the same locus on the same chromosome (20).

The contact system. Genes for factor XI and prekallikrein were not found in puffer fish, the best match found on back-searching in both cases being a mosquito protease. In line with these observations, no apple domains were found to be associated with serine proteases, although several such domains were found on a relatively short scaffold (scaffold 11138) that could conceivably be associated with some protease on another scaffold.

A gene for factor XII could not be found, the closest candidate on back-searching in this case being salmon trypsin. A search for scaffolds containing both kringles and EGF domains did not reveal any proteases with domainal arrangements similar to what occurs in mammalian factor XII.

Factors V and VIII. The genes for factors V and VIII, themselves paralogs descended from ceruloplasmin (15), were readily identified. These very large proteins are comprised of three A domains and two carboxyl-terminal discoidin domains (denoted C in Fig. 3). The B region between the second and third A domains varies greatly between the puffer fish and human proteins, as it does between factors V and VIII from any species. This region has been found to be expendable in human factor VIII expression systems (23).

Fibrinogen genes. The β- and γ-chains of puffer fish fibrinogen are 58% and 52% identical with their human counterparts. In the case of the β-chain gene, the puffer fish has four additional
introns, all quite short, besides the ones in common with the human gene. In the γ-chain gene, all introns were found at approximately the same positions as occur in the human gene except for one missing from the puffer fish that marks the splice site near the carboxy terminus of mammalian γ-chains. This intron, which was previously found to be absent from the lamprey γ-chain gene (24), is what in mammals leads to a minor form of fibrinogen with γ' chains that serve as binding sites for factor XIII (25).

The fibrinogen α-chain in puffer fish has a similar arrangement to what is found in chickens, none of the central repeats found in mammalian or lamprey α chains being present. Additionally, several large deletion events are apparent in the αC domain. The alternatively spliced αCe domain characteristic of the minor extended form of fibrinogen known as fibrinogen-420 is encoded in the downstream region of the gene, just as occurs in birds and mammals (26). In lamprey, the minor form of the α-chain is encoded in an entirely separate gene denoted α2 (27).

Factor XIII. The A chain of factor XIII was obvious (48% identical to human factor XIII); a second somewhat truncated version of the protein was found on another scaffold and was only 43% identical. A gene for the factor XIII accessory B chain was not found. In mammals, the B chain is composed of 10 "sushi" domains (also known as complement-control modules or β2-glycoprotein domains). These domains are quite common in mammalian proteins where they occur in numerous components of the complement system. The domains are also common in the puffer fish genome, but no appropriate string of ten was found that corresponds to factor XIII B. The best match occurred between a cluster of a dozen or more sushi domains on scaffold 310 that on back-searching identified a mouse polydomain protein containing 18 such domains.

Fibrinolytic proteases. Orthologous genes for plasminogen, t-PA (tissue plasminogen activator) and u-PA (urokinase-type plasminogen activator), were readily identified. The plasminogen gene actually spans three different scaffolds. The main body of the protein, including kringle 4 and 5, is encoded on scaffold 9368 (4148 nt), but the carboxy-terminal region is encoded at the 5′-end of scaffold 1457 (160,920 nt), and the amino terminus, including a PAN domain and kringle 1–3, at the 3′-end of scaffold 1979 (48,368 nt).

The t-PA gene has the expected FN-1 domain (Fig. 4A), an EGF domain, and two kringle. In the case of u-PA, the terminal EGF domain seems to be fragmented, but the single kringle is intact. In all cases, introns separate the peripheral domains from each other and the serine protease portions of the genes.

Serpins. A gene for antithrombin III was found whose putative protein is 55% identical with the human protein. The protein has two introns not found in the human gene, as well as five that they have in common.

The search for plasminogen activator inhibitor 1 (PAI-1) resulted in a failed back-search. The initial search of the puffer fish genome with human PAI-1 as a probe identified a gene on scaffold 1754 whose gene product was 44% identical. When that sequence was back-searched against GenBank, however, the best hit, at 59% identity, was leucocyte elastase inhibitor. Contrarily, when the human α2 antiplasmin sequence was searched against puffer fish, a best match was found with a gene whose protein product was only 34% identical. Nonetheless, the back-search against GenBank turned up α2 antiplasmin as the best hit. For the moment, we have cautiously labeled this as an ortholog, even though the resemblance is lower than expected.

Other inhibitors. Two genes for tissue factor inhibitor (TFI) were identified, just as is observed in humans. The best match among these was 44% identical. As for other inhibitors, a gene for the carboxypeptidase known as thrombin-activated fibrinolysis inhibitor was found, the putative protein sequence of which is 45% identical with its human counterpart.

Two genes that correspond to thrombomodulin were found adjacent to each other on puffer fish scaffold 195; an analogous situation of two adjacent genes occurs in humans, in which case one of the genes is thrombomodulin and the other is thought to be an accessory protein for complement factor C1Q (28). As is also the case in humans, neither of the two puffer fish genes has introns.

Peripheral domain inventory. All told, 19 GLA domains were found in the puffer fish genome (Table 2), 11 of which were associated with the vitamin K-dependent serine protease clotting factors. Two others were parts of protein S and its homolog. None of the remaining six were associated with EGF domains or serine protease domains.

The total number of kringle found was 34, of which 5 were in plasminogen, 2 were in t-PA, 1 was in u-PA, and 2 were in prothrombin, leaving about two dozen unaccounted for. Many of these are likely associated with nonclotting proteins like hepatocyte growth factor (29), but we did not pursue the matter further.

Only two FN1 domains were found in the puffer fish genome (Fig. 4A), but at least 19 FN2 domains are present (Table 2).

Table 2. Some domains in puffer fish and sea squirt genomes*

<table>
<thead>
<tr>
<th>Domain type</th>
<th>Puffer fish</th>
<th>Sea squirt</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLA</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>Kringle</td>
<td>34</td>
<td>50</td>
</tr>
<tr>
<td>Apple (PAN)</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>F5/B-A</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>F5/B-C (discoidin)</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>FN1</td>
<td>2†</td>
<td>0</td>
</tr>
<tr>
<td>FN2</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>FN3</td>
<td>&gt;100</td>
<td>&gt;100</td>
</tr>
<tr>
<td>EGF</td>
<td>&gt;100</td>
<td>&gt;75</td>
</tr>
</tbody>
</table>

*Domains were identified with TBLASTN using cutoff E = 10⁻³.
†Cutoff E⁻¹; see also Fig. 4A.
None of the 27 apple domains found in the puffer fish genome were associated with proteases, although, as noted above in some cases, the scaffolds were small, and associated proteins may not have been recovered in the assembly process. We also found 12 F5/8-A domains and 27 F5/8-C (discoidin) domains (Table 2).

**Sea Squirt Genes.** None of the 26 coagulation factors turned up a convincing case for an orthologous gene. Nonetheless, almost all of the constituent domains that are generally associated with these factors were found in one context or other, including GLA, sushi, FN2, FN3, EGF, and F5/8-A and -C domains. The two exceptions were the FN1 and apple (PAN) domains, neither of which were found.

Only four GLA domains were found to be encoded in the sea squirt genome. None of these were associated with kringles, making it unlikely that there is a prothrombin-like gene. On the other hand, all of the GLAs are associated with multiple EGF domains, but none of these is near a serine protease domain. In one case, the GLA domain is near an EGF domain but in the wrong order; i.e., the EGF is on the amino-terminal side of the GLA.

All told, 50 kringles were found in sea squirt genome, including eight situations where multiple kringles are found with serine proteases. These included four sets of two, three sets of four, and one set of five. However, neither plasmin, t-PA, nor u-PA survived the back-search process.

Several multikringle serine proteases were found that might be construed as “plasminogen-like.” That the best of these was not an authentic ortholog was attested to by the lack of a terminal PAN domain (13), the first 120 residues being instead more like a domain found in thrombospondins (30). Moreover, the putative serine protease portion failed the back-search test, being more similar to a trypsin from a sponge than to any plasmin; indeed, plasmin was not among the top 100 hits.

A number of sequences were identified that are homologous to the carboxyl-terminal domains of fibrinogen, not unexpected because these domains are widespread in animals (31); however, no full-length genes were found with the potential for the constituent coiled coils that are hallmarks of fibrinogen.

One intriguing situation involved a gene on scaffold 87 conceivably related to factors V and VIII, which, as noted above, are descended from ceruloplasmin (15). A putative protein was identified that has both A domains and a C domain. The A domains are much more similar to ceruloplasmin than they are to either factor V or factor VIII, however, even though the sea squirt has a transglutaminase, but there was no evidence of the signal peptide required for a circulating factor XIII, and it is likely an ordinary tissue transglutaminase. Similarly, a number of serpins are encoded in the genome, but they could not be matched with antithrombin. The sea squirt also has a carboxypeptidase paralogous to thrombin-activated fibrinolysis inhibitor, as well as numerous domains that are similar to those found in tissue factor, tissue factor inhibitor, and thrombomodulin.

**Discussion**

**Late-Appearing Factors.** The presence of many genes for mammalian clotting proteins in the puffer fish genome was anticipated, a number of them having been isolated and/or cloned from more primitive fish in the past. For example, prothrombin has been isolated from lamprey blood plasma (2) and cloned from the hagfish (32). Moreover, the principal ingredient of blood clots, fibrinogen, has been fully characterized (sequenced, cloned, and fragments crystallized) from lamprey (1–3, 33). As such, the events involving the thrombin-catalyzed conversion of fibrinogen to fibrin and its subsequent cross-linking and lysis were not an issue. The exact pathways of thrombin generation were still unclear, however, and it remained uncertain whether all of the vitamin K-dependent factors are present in lower vertebrates, although it should be noted that factor VII, as well as prothrombin, has been identified in zebrafish (34, 35).

It was expected that some components of the “contact phase” system (factors XI and XII, and prekallikrein) would be missing from lower animals because amino acid sequence comparisons showed that factor XI and prekallikrein are so similar that the gene duplication leading to them must have been very recent (4). Nonetheless, the complete absence of the contact phase system from the puffer fish genome is noteworthy. Of course it cannot be ruled out that one or all of the absent genes do not reside among the 5% of DNA sequence data that are not yet available from the puffer fish. This absence aside, it is remarkable how stable the rest of the coagulation scheme has been over the course of the last 400 million years.

**Assembling the Scheme.** It is thought that 50–100 million years separate the appearances of urochordates (which include the sea squirt) and vertebrates. During that time the machinery for thrombin-catalyzed fibrin formation had to be concocted by gene duplication and the shuffling about of key modular domains. The relative times of duplicative events can be estimated by various means, the most obvious being the presence or absence of a gene in earlier diverging organisms, although it must be kept in mind that lineages may lose genes. Another way to gauge events is from the relative positions of various gene products on phylogenetic trees, earlier branching implying earlier appearance. In this regard, (pro)thrombin invariably appears lower on the phylogenetic trees than do the other vitamin K-dependent factors (Fig. 2).

The order of events can also be inferred by considering the most parsimonious route to assembling the various clusters of peripheral domains. Nine of the proteases under discussion can be accounted for by six domain-swapping events (Fig. 5). Indeed, the presence of a multiple-kringle protease in the sea squirt genome provides a reasonable model for a step-by-step parallel evolution of the clotting and lysis systems. It should be noted that
a serine protease with only one kringle has been found in the ascidian *Herdmania momus* (36). Although numerous scenarios have been offered in the past about how modular exchange was involved in generating these schemes (refs. 4, 12, and 37–41, *inter alia*), the new genomic data now provide a realistic set of starting materials.

The timing of duplicative events can also be approximated from ortholog–paralog comparisons. As an example, human and puffer fish factor V are 41% identical, and human and puffer fish factor VIII are 42% identical (not counting the variable B regions). On the average, the two factors themselves (in this region) are 38% identical, implying that the gene duplication that led to them occurred only a relatively short while before the common ancestor of fish and mammals. The difference is so small (42% vs. 38%) that it may turn out that the earlier diverging jawless fish will have only the preduplication gene. A genome study devoted to the lamprey or hagfish would settle the point.

We are grateful to Da-Fei Feng for his help and encouragement during the course of this project. This work was supported by National Institutes of Health Grant HL-26873.


7532 | www.pnas.org/cgi/doi/10.1073/pnas.0932632100
FTM1 >Fugu thrombomodulin on Scaffold_195  
(sequence length=542)

HTM1 >gi|1070535|thrombomodulin precursor [validated] - human  
(sequence length=575)

Number of matches = 202
Fraction of identities per length = 0.372694

M N D V T R L I A I V C A L L W G R T G G V K P N N I Y C
M L G V L V L G A L A L A G L G F P A P A E P Q P G S Q C_

I G N Q C F T V S K I L G D F S A A Q S E C Q D R G G N L M
V E H D C F A L Y P G P A T F L N A S Q I C D G L R G H L M_

T V R S S V A H D V L L M L L G D A S S R F W I G L
T V R S S V A A D V I S L L L N G D G G V G R R R L W I G L_


N W P T T F N S S C S S H R C V S V S P A D G F
R W A R D L N G A P L C G P L C V A V S A A E A T V P S_

K W T R E P C D T R T D G F L C E Y S F N E T C N S L E
E P I W E E Q Q C E P K A D G F L C E F H F P A T C R P A_

A G A G E T L T Y S T P I G F E G E D L L S L L P
V E P G A A A A A A V S I T Y G T P F A A R G A D P Q A L P V_

G S T A V R M P T E T K Y I C F S G Q W T Q
G S S A A A V A P L G L Q L M C T A P P G A V Q G H W A R E A_

A P W N C D V L G G G C E Y K T C T Q D P E K M P L C Y C P
P G A W D C S V E N G G C E H A C N A I P G A P R C Q C P_

R G K T I N P E N E V T C E T Q E D P C A S L R C A H I C
A G A A L Q A D G R S C T A S A T Q S C N D L C E H F C_

Y E N G S F A C T C R E G F K L A A D G R S C V D F
V P N P D Q P G S Y S Y M C E T G Y R L A A D Q H R C E D V_

N D C T D P R Q C P G E N S R C V N T A G G F Q C V K D
D D C I L E P S P C P Q R C V N T Q G G F E C H C Y P

G Y R Y K G G V C V D V N E C T S A P C E H M C D N L P
N Y D L V D G E C V E P V D P C F R A N C E Y Q C Q P L N Q_

G S Y V C S C Y P G Y K E D P E E P H R C K L Y C G E E E C
T S Y L C V C A E G F A P I P H E P H R C Q M F C N Q T A C_

L A E C D P N D R Y Q C F C P E G Y I L E R Q Q D A V C L
P A D C D P N T Q A S C E C P E G Y I L D D G F I C T

D I D E C S S F Y C D Q D C E N T Y G S Y V C I C S P G Y
D I D E C E N G G F C S G V C H N L P G T F E C I C G P D S
K L V D N Y K C V K N E V G P D E G S G A S S T P P S L
FTFI >fugu TFPI on scaf611, from tbn
   (sequence length=288)
TFPI >gi|125932|sp|human Tissue factor pathway inhibitor (TFPI)
   (sequence length=304)
Number of matches = 102
Fraction of identities per length = 0.354167

FTM2 >fugu thrombomodulin candidata2 on Scaffold195
   (sequence length=472)
HTM1 >gi|1070535|thrombomodulin precursor [validated] - human
FTFI >fugu tissue factor inhibitor on scaf1267
(sequence length=241)
TFI2 >gi|5730091| tissue factor pathway inhibitor 2 [Homo sapiens]
(sequence length=235)

Number of matches = 103
Fraction of identities per length = 0.438298

FALF >part of the genscan of alpha on Scaffold_3291
(sequence length=697)
FALC >gi|971185|gb|AAB60686.1| fibrinogen alpha-E subunit [Gallus gallus
(sequence length=741)

Number of matches = 258
Fraction of identities per length = 0.370158

FMQA QF
T Q N D T L L R V E L E D W D G N A A Y A E Y I V Q V G T
E E E G F P L H V S G Y S G E G G D A L V L S G S D M A P S
L S H D G M K F S T F D K D N D K W D Q N C A E V Y G G G
Y T S H A Q M Q F S T F D R D O D H W E E S C A E V Y G G G
W W Y N R C Q S A N L N G V Y Y R G S Y N P E K N P A
W W Y N S C Q A A N L N G I Y Y P G G H Y D P R Y N V P Y E
V H N G V V W S T F N S S Y S L K A
I E N G V V I P F R A S D Y S L K V V R M K I R P L E T L

FAPL >fugu antiplasmin on scaf1092
(sequence length=460)
HAPL >gi|178751|gb|AAA35543.1| alpha-2-antiplasmin precursor
(sequence length=488)

Number of matches = 159
Fraction of identities per length = 0.345652

M A L W F R R L S L L N R K \ K P E N N S A T K V P A A A N T
L W \ G L L V L S W S C L Q G P C S V F S P V / \ S A M E
S Q P D S S E D G R N E D Y C L I G R S L E S R E A I A
P L G R Q \ L T S C P N Q E Q V S P L T L L K L G N Q / \ E P G G
A A I Q K L G V Q L L Q N L E A T P E Q P N I I I S P L S / \ A
Q T A L K _ S P P G V C S R D P T E Q T H R L A R A M M A_
F T A S P L S F V _ S I F P K F M A A W R C P H A L I
F T A D L F S L V A Q T S C P N L I L S P L S V A L A L S
A L F L A G A V N E T R E L L M H H L H E R A L P C Y H E S
H L A L / \ G A Q N H T L Q R L O Q V L H A G S G P C L P H L
L H N I L A G L R K N D L Q I A T Q I F L R Q \ / G F Q P K Q D
L S R L C Q D L G P R F L A R M Y L Q K / \ G F P I K E D_
F V N K S R H L Y G S P A E L K S _ L Q Q I N D W
F L E Q S E Q L F G A K P V S L T G K Q E D D L A N I N Q W_
V Y N A T N G K M P Q F L S A L P L N V L V M L I N A V H F
V K E A T E G K I Q E F L S G L P E D T V L L L L L N A I H F_
K / \ G W V A R F D F R F T S R G A F Y L D D N N M I D V E V
Q / \ G F W R N K F D P S T Q R D S F H L D E Q F T V P V E M
M E D A K H P L S L F I D N E M D A Q / \ V M Q V A R F R F R K
M Q A R T Y P L R W F L L E Q P E I Q V / \ A D F P F K N
R L P K E R A V Q V K V P K F K L E Y S Q E L Q E V F T K I \ /
PLVWERPKVRLPKLYLKHQMDLVATLSQ/

GGLGEIFSRSPLTAIEADGPLLVSVSVMKST
GLQELFAQPDRLGRSEQSLVSGVQHQRST
MEINEEGAAATTVISRASSPVFHTQ
LELSEVGVEAAATSIAMSRMRSSFSVNR
PFFFAVMDDTTEVPIMFGVVNNPNPGAP
PFLFIIFEDTTGLPLFVGSRNPNPSPAPRE
VMQTGDKVGFPIDKSM
LKEQQDSPGNKDFQLSLSKLGFPRGDKLFGPD
TRFEGPPKLKLVPPMEEYPQFGSPPK

FUAT >fugu antithrombin-III scandidate on Scaffold_1063
(sequence length=437)

HSAT >gi|113936|sp|P01008|ANT3_HUMAN Antithrombin-III precursor (ATIII)
(sequence length=464)

Number of matches = 241
Fraction of identities per length = 0.551487

MPASDWWLLLLASLHVVS
MYSNVICTVTSGKRKVYLLSLLLIGFWDCV

ADVLDICGAKPRDLALEPRCITYSPDP
TCHGSPVDICTAKPRDIPMNPMCIYRSPEK

EAPPEPLTTHPVPGSTNPVRWELSKANARF
KATEDEGSEQKIDPEATNRRWELSKANRSF

AMSLYKQVASSRGPNINIFMSSPISISTAF
ATTFYQHLADSNNDNINFLSPLSISTAF

MTKLGACNQTELQLMR//VFEFDTIKEKTSDDQ
M_TKLGACNQTELQLME//VFKFDTIKESKTSDDQ

VHFFFAKLNCRLYRKDKSNELVSANRLF
IHFFFAKLNCRLYRKANKSKLVSANRLF

DKSLAFDQTYQNISETVYGAKLLPLDFK//DD
DKSLTFNETYQIELVYGAKLQLPQLDFK//EN

PEKARVTINWISNKTENLIQDTLPAGG//GA
EEQSRAAINTKWSNKTEGRTDVIPEAII

HWWKNIKFDKDNVYVSEFHSSQTTSRL
ELTVLNVNTIYF//GLWKSFKSFENTRKE

GQHDVPGPSAPFKR//PFYKHFPEDQV
FYKADGESCSCASMYQEGKFRYRRVAEGT

QLLEMPYRGDITMVILPSQGTALSQ//VEE
FUPC >fugu protein C on scaffold 8062
(sequence length=448)
HSPC >gi|21707771|gb|AAH34377.1| human protein
(sequence length=461)
Number of matches = 204
Fraction of identities per length = 0.455357

MLRPVLCASVAVWWSASVGLGLS/\VFSN
MWQLTSSL_LLVFWATGIGSTPAPL/\DS_VFS_S
APDAHMLLRSR_KRANSFLEELKPPSMEREC
SERAHQVLRIRKRANSFLEELRHSLSERE_C
V E E N C D F E E A R E I F Q T R E A T V R V A V N A C / \ I L
I E E I C D F E E A K E I F Q N V D D T L A F W S K H V / \ D G
GNHELLPDGNQCDNNMVCNVTCDKVYQAYA
DQCLVLPHELPCASLCCGHGTCLUDGISFS
CSCNHYGEGRYCDQ//PLTATNCSDLNGNCDH
C D C R S G W G R F C Q R / \EVSFLNCSDLNGC_T H
ECTDGDAGGLTRRGCVCNVYNLQDDSRTRCP
YCLEEVGWRRCSCAPGYKLDLQLCHP_
K/\GPSSCGQLLIGRSSTYTKSJDG
A/\VKFPCCGRPKRMEEKKS_HLKRDTEDQEDQ
LLPWMDGVGEVGGKGESPWO/\VLVLANVGGKHF
VDPRLIDGKMTRGDSPWQ/\VVLDSLKKKLA
CGGVLEDIESWVLTAHCLEDSLTFRVRGLGD
CGAVLIIHPWSWVLTAHCMDESKKKLLVRL/\GE
YERLRAEGTVEVTKLVTKTFKHPKYNRSVD
YDLRRWKEWLDDLIDKEVFVHPNYSKSTTD_
NDISLLRL/\ETPAPLSDYIVPVCLPGRLAQ
NDIALLHL_ AQPATLSQTIVPICLPDSGLAEE
RVLNKNTMTVVSGLGW_ KENLESSRFS
RELNOAQGQETLVTGWGYHSSREKEAKRNRT
SALNVKPVPLVDTDTGCRQMYYNITSNMLCVFLNFLIKIPVPHNECSEVMSNMVSENMLC
AGIVGQKMDACEGDSGGPMVTLYRDTWWFLV
AGILGDRQDACEGDSGGPMSFHTWWFLV
GLVSWGEGCGNVEKLGITYKVSNYIDWINKG_LVSWGECGGLLHNYGVYTKVSRYLDDLWHG
VREDWDTSIPERVQRP
HIRDKEAPQKSWAP

PSFU >fugu protein S on Scaffold_2356 646_aa
(sequence length=646)
PSHU >gi|190442|gb|AAA60180.1| protein S alpha [Homo sapiens]
(sequence length=650)

Number of matches = 331
Fraction of identities per length = 0.512384

|LSPSTASQQLRRHRANSLFEESKPGNLER|
|LSKQQAASQLVRRKARANSLLEETKQGNLER|
|ECIEELCNKREAMERIFENQPET/\ EYFYPKYYV|
|ECIEELCNKREAMERVFENDPET/\ DYFYPKYL|
|/\VCLGSHHRVGI_ GNHOPGISDLRTCVT/\ E|
|VCLRFSQQTGFLTAARQSTNAYPDRLRSCVN/\ A|
|INNQCSPYPCYKEGSLRCVDQASFTCVCK|
|IPDQCSPCLPNCEDGYMSCDKDGKASFTCTCK|
|PGWKQKCRE/\ DIDECCLDP EFPAGCNKCN|
|PGWQGKEKCEF/\ DINECKDPSNIGCSCQICD|
|NIPGSFQCQCESGHDYFL NQITCV/\ DVEDCQ|
|NTPGSYHCSCKNGFVMLSNKDKC/\ DVEDCQ|
|LYPSICKEPARCVNSPGMYECRC PKGF RYNN|
|LKSNI_C_ GTAVCKNIIPGDFECECPEGYRN|
|FTSKTCS/\ DVEDCEMSVCDGICIINTVGSYECLKSRSCE/\ DIDESENMCAQLCVNYPGGYTC|
|HCDGRLGLRLAEVRQC RIPVCVDLYDHHK|
|YCDGGFKLQAQDQKSC/\ VSVCLPLNLDT|
|HSEMLYLGEOFSLPAFMFLRYRLPENTK/\ FA|
|KYELLYLAEOQFAGV VLYLKFRLPEISR/\ F_S|
|AEFDFTFDPDEGVVLYAES SQGSWFMLGL|
A_E_F_D_F_R_T_Y_D_S_E_G_V_I_L_Y_A_E_S_I_D_H_S_A_W_L_L_I_A_L_
R_G_G_H_I_E_V_Q_F_K_N_Q_H_T_F_K_L_T_S_G_G_K_A_I_N_D_G
R_G_G_K_I_E_V_Q_L_K_N_E_T_S_K_I_T_T_G_G_D_V_I_N_N_G_L_W_N
T_\_/ I_S_V_D_E_L_E_S_S_I_S_V_K_I_S_K_E_A_V_M_S_I_N_S_P_Q_S_L_F
M_\_/ V_S_V_E_E_L_E_H_S_I_S_I_K_I_A_K_E_A_V_M_D_I_N_K_P_G_P_L_F_
T_A_V_N_G_K_V_E_T_K_V_Y_I_A_G_L_P_E_R_A_D_T_\_I_K_P_\_/ I_N_P
K_P_E_N_G_L_L_E_T_K_V_Y_F_A_G_F_P_R_K_V_E_S_E_L_I_K_P_\_/ I_N_P_
R_L_D_G_C_I_R_G_W_N_L_M_N_Q_G_A_S_R_V_K_E_V_I_Q_E_L_K_S_K_Q
R_L_D_G_C_I_R_S_W_N_L_M_K_Q_G_A_S_G_I_K_E_I_I_Q_E_K_Q_N_K_H
C_F_I_S_V_E_K_G_S_F_F_S_G_M_G_L_A_S_F_N_V_D_Y_\_/ S_D_S_G
C_L_V_T_E_K_G_S_Y_Y_P_G_S_G_I_A_Q_F_H_I_D_Y_\_/ N_N_V_S_S_A_E
S_W_S_V_D_I_E_M_N_I_R_P_S_S_S_T_G_V_I_F_A_L_V_S_N_D_T_V_P
G_W_H_V_N_V_T_L_N_I_R_P_S_T_G_T_G_V_M_L_A_L_V_S_G_N_T_V_P_
L_S_I_A_V_V_T_Q_G_E_G_E_A_\_/ N_L_Q_V_F_L_G_G_V_S_V_AT_L_D_S
F_A_V_S_L_V_D_S_T_S_E_K_S_Q_\_/ D_I_L_S_V_E_N_T_V_I_Y_R_I_Q_A
L_M_L_C_Y_P_E_R_L_T_V_S_L_K_I_T_P_A_A_V_Q_V_S_G_N_S_S_T_V_T
L_S_L_C_S_D_Q_Q_S_H_L_E_F_R_V_N_R_R_N_N_L_E_L_S_\_T_F_L_K_I_E
Y_V_T_S_E_S_L_Q_E_A_L_E_H_L_N_A_T_M_Q_N_P_L_T_T_Y_I_G_G_I_P
T_I_S_H_E_D_L_Q_R_Q_L_A_V_L_D_K_A_M_K_A_K_V_A_T_Y_L_G_C_L_P_
\_/ D_D_I_P_L_P_A_T_P_V_T_A_Y_Y_H_G_C_M_D_I_S_V_N_G_Q_Q_L_D_F_D
\_/ D_V_P_F_S_A_T_P_V_N_A_F_Y_N_G_C_M_E_V_N_I_N_G_V_Q_L_D_L_D_
E_A_L_S_K_H_N_S_I_K_S_H_S_C_P_P_V_S_A_P_D_R_Q_G_D_V_L_Q_P_P
E_A_I_S_K_H_N_D_I_R_A_H_S_C_P_S_V_W_K_K_T_K_N_S
A_E

FUPT >fugu prothrombin on scaffold 403
(sequence length=618)
HSPT >gi|339641|gb|AAC63054.1| prothrombin [Homo sapiens
(sequence length=622)

Number of matches = 323
Fraction of identities per length = 0.522654
M A G P T V K P L T V L L L L F L L H S C L A N H \_/ V F L S
M_A_H_V_R_G_L_Q_L_P_G_C_L_A_L_A_A_L_C_S_L_V_H_S_Q_H_\_/ V_F_L_A
S R S A S Q V L V R S R R A N Q M F E E I K A \_/ G N L E R E C
P Q Q A R_S_L_L_Q_R_V_R_A_N_T_F_L_E_E_V_R_K G_N_L_E_R_E_C_
M E E K C N Q E E_A_R_ E_V_F_E_Q P_D_T_T_\_/ E_A_F_W_K_K_Y_L_\_/ D C
V_E_E_T_C_S_Y_E_E_A_F_E_A_L_E_S_S_T_A_T_\_/ D_V_F_W_A_K_Y_T A_C_
N G N Q_L_P_R_T_E_A_N_I_I_A_F_K_E_C_L_D_\_/ G_F_C_I_S_G_R_G_L_N
E_T_\_/ A_R_T_P_R_D_K L_A_A_C_L_E_\_/ G_N_C_A_E_G_L_G_T_N_
Y d G V N I S K G I Q C Q H W K H S F P H P I M R \ / E Y N
Y R G H V N I T R S G I E C O L W R S R Y P H K P E I \ / N S
A S E P D S I L Q E N F C R N P N S P D G P W C F T T D P
T T H P G A D L Q E N F C R N P D S T T G P W C Y T T D P_
T V Q K E T C R V P I C G \ / V F L I P G E A F V P P T P T P P K
T V R R Q E C S I P C G _ \ / Q D Q V T V A M T P R S E G S S V
H F K T I Q S S C I S N Y G V D Y V G D L D V S A K G H A C
N L S P P L E Q C V P D R G Q Q Y Q G R L A V T T H G L P C_
L M W S S P E A V T L S Q N K E F D P D I N L L C N K C R N
L A W A S A Q A K A L S K H Q D F N S A V Q L V E N F C R N_
P D K D P E G P W C Y V N A S G K V I V D Y C D L P V C \ / E D
F N P R T F G E G E D \ / E C G R R P L F E Q K N K K D A S E D
F N P R T F G S G E A \ / D C G L R P L F E K K S L E D K T E R
E L L Q S Y R E K R I V G G D E A E V A S A P W \ / Q V M L Y K
E L L E S Y I D G R I V G S D A E I G M S P W \ / Q V M L F R
N K N F S A S D I L V R L G K H N R A K \ / F E Q G I E K I M V
D K N F T E N D L L V R I G K H S R T R \ / Y E R N I E K I S M

TAFI > Fugu thrombin-activatable fibrinolysis inhibitor on scaf123 (sequence length=421)
TAFI > gi|4503005| thrombin-activatable fibrinolysis inhibitor (TAFI) (sequence length=423)
(sequence length=467)

HSFX >gi|20336663|gb|AAM19347.1|AF503510_1 coagulation factor X [Homo sapiens
(sequence length=488)

Number of matches = 207
Fraction of identities per length = 0.443255
Number of matches = 344
Fraction of identities per length = 0.469945

M S D P A P V V D P S P A P A A P S G P R P K V T N R G R T
M_SET_S_R_T_A_F_G_G_R_R
A V A V A S N S E S T E V P E F E S F V A L S P R G F P P
A_V_P P N N S N A A E D D L P T V E_L Q G V P R G_V N
L_T D Y_/L D I V A V D M M S Q S D G V N K Q Q H R T M F Y N
L_Q_E_F_/L N V T S V_H L F K E R W D T N K V D H H T D K Y E
S_N Y_L I V R R G_Q_E_F_H_V K_L T F N R P Y N P_E_D K F A
N_N K_L I_V_R_R G_Q_S_F_Y_V_Q I D_F_S_R_P_Y_D_P_R R D L_F_R
L_E_F_V I G_/A N P_D Y S K G T Y I_P V_F_P N_K E_R Q_S R W
V_E_Y V_I G_/R Y P_Q E_N K_G_T_Y I_P V_P I V S E_L Q S G K W_
A G_R I A D_S S N N D V T V G I T P L A N C I V G K Y H M_Y
G_A K_I_V_M R E_D R S V R L S I_Q_S S P K C I_V K_F R M Y_
I_A V M T P F G I R R T R K D P V R D L Y I L F N P W_/S P_A
V_A V_W T P_Y G V L R T_S R N P E T D T Y I L F N P W_CE_/D
D_D V F L D D_E P E R Q E C V M N E M G I I Y H G A Y D D_I
D_A_V Y L_D N E K E_R E_E Y V L N D I G I_V F Y G E V N D_I_
A E R Q W N Y G Q_/F N Y G V L D A C L Y I M D R S E M P I_T
K_T R_S W_S Y_G_Q_/F E D G I_L D T C L_Y V M D R A Q M D L_S
N_R G D P I K V T R K A S A M_/L N_S R D D D G V L V G N W_S
G D Y T Y G V A P T S W T G S T E I L L T Y A S S R M P V_S
N I Y A G Y G_V P S A W T G S V D I L L E Y R S S E N P_V_R
FUF8 >fugu f8 candidate on scaf2929  
(sequence length=1583)  
HSF8 >gi|66384|pir||EZHU coagulation factor VIII precursor [validated] - huma  
(sequence length=2351)  

Number of matches = 658  
Fraction of identities per length = 0.415666  

M R T D       T L L L P L P                       L P L L L  
M_Q I E L S T C F F L C L_L_R F C F S A T R Y Y L G A V E  
L V A A F C G G D A Q Q       P P A A V R Q Y Y I  
L_S W D Y M Q S D_L G E L P V D A R / \ F P_P R V P K S F P F N  
A A V E I G W D Y I H V E D       G D P A S E Q R / \ G I  
T S V Y K K T L F V E T D H L F N I A K P R P P W M / \ G L  
Q G P V I V A Q A G E T V L V H F K N L A S Q P Y S I S P V  
L G_P T I_O Q A E V Y D T V_I T L K N M A_S H_P V S_L H A V_  
G I S Y W K H S E G / \ A G Y D D S T A G H E K E D D A V Y P G  
G_V S Y W K A S E G / \ A E Y D D_Q T S Q R E K E D D K V F P_G_  
G Y Y E Y V W D I S P K D G P T G S D P E C L T Y S Y S S Q  
G_S H T Y V_W Q V L K E N G P_M A_S D_P_L C L_T Y_S L S_H  
V D P V R D V N S G L I G A L I C K / \ M S A F T D E G Q R R  
V_D L V_K D_L N S G_L I G A L I C K V C R / \ E G S L A K E K T Q T  
N Q A F V L L F V V F D E         S K S W Y G E V G E R K S R D K  
L H K F I L_L F A V_F_D_E / \ G K_S W H S E T K N S L M Q D R D
NGYVAESLPGLLVAGHRQVRWHLNLNVGSDG
NGYIMDTLPGLVMQAODQRIRWYLLSMSGSNE
EYHAVFHFGLPFTVHAKKEHRMGMVYNLFPP\G
NIHISIFHSGHVFTVVKKEEYKMALYNLYP\G_
VFGETVEMPPTVTWLVECTVGEQLAGMR
VFETVEMPSPKAGVRWECLIGEHLMHAGMS
AKLLVYNP\QCPSRPLGMKSGRIGDSQIKASD
TLFLVYN\KCQTPPLGMASSGHIRDFQITASG
YIGNIILFISPVRSQYG/\QPAPKLARLHYSGSINAWSTKFPFSWI
KRNLGYRVQTOQGVRSNLRNNYITAF
KVDALLAPMIHGKTQRQAKFSSLYISQF_
TVSYSUDQETWSTYRGs\GSSRSSSSSTSTA\KV
IIIMYSLDGKKKWQTYRG\NSTGTLM/V_
FNGNLDNSRVKNNPFPFPFPVARYIRIHPLY
FGNVDSSGIKHNIFNPIARYIRLHPETH_
YNQRPALMELLLGCDLN/SCSLPLGLQDRRI
YSISRSTLRMELMCGLDNL/SCSMPLGMESSKAI_
PDESFVASSSYWSLLRSTWPSLARLHQEGS
SDAOQITASSYFTNMFATWSSPKARLHLQGR
ANAWRPK/\NNNPHEWLQVDLGGVRITGTGVVT
SNAWRPQ/\VNPKEWLOQDFQKTMKVTGVTT_
QGARSLTKMMVETFSTISRDGQAWS\VL
QGVSLLTSMEYVEKFLISSQDQGHWTLFF
EGSSQREKIFQGNNDSDDEEALTIFDAPLFG
QNGKVK/VFQGNGQDSFTPVNVSLDPPLLTT
RYIRIHIPGWINDIARLRELVEGLCDAQALR
RYLRIHPQSWVHQIAIARMEVLGCEAQLDLY

GAFU >fugu gamma on scaffold 39
(sequence length=442)
GAHU >gi|4503715 fibrinogen, gamma chain
(sequence length=437)

Number of matches = 223
Fraction of identities per length = 0.510297

MLLAIRTVPNKLCSWHLHQPETLWT\VTE
MSWSLHPRNLILFYAYALLFLSSSTCVA\\Y
VAT
RSFLSCCTFLGSPTS\GMYCPTKCGVADYML
RDNC_CILDERFGSYCPTTCGIADFLS
FPAI >Fugu plasminogen activator inhibitor on scaf1754
(sequence length=406)
HPAI >gi|189545|gb|AAA36413.1| plasminogen activator inhibitor
(sequence length=415)

Number of matches = 181
Fraction of identities per length = 0.445813

M A A I S G S N T A F A L E L L R T L S Q G N P S G N I F V
M E D L C V A N T L F A L N L F K H L A K A S P T Q N L F L
S P L S I S S A L A M V Y L G A K G E T A A Q M A Q V M D S
S P W S I S S T M A M V Y M G S R G S T E D Q M A K V L Q F
N     G R I E S L Q W D K S D F     P C     L Q A L       S F
N_E V G_A N A V T P M T P E N F_T S C_G F M Q_Q I Q K G S_Y
S               S G K D V H A D F Q T L N G E I N S P S A S
P D A I L Q A Q A A D K I H_S S F_R S L_S S A I N A S T G D
Y T L K L A N R L Y G E T A N F L S V G S G F T P I T N C
Y_L L E S V N K_L F_G E T K S A_S F_R E E Y I R L C_
Q K Y Y H A D L K A I D F I G A T E E C R A E I N S W V E E
Q_K Y Y_S E P Q A_V D_F L E C A E E A R K I N S W V K T
Q T E N K I K D L L K P G T V S T M T R L A L V V N A I Y F K
Y Q M K K L P Y N Y I P E H G V Q I L E L Y P Y V E E E L S M
Y_L R E K L N I G Y I E D L K A Q I L E L P Y_A G D V S M
F I L L P E E T T D G P S P L L K L E N E L T R E K L D E W
F_L_L_L_P_D E I A D V S T G L E L E S E I T Y D K L N K W
T N R E N M D V H S E V L V H L P K F K L E E D Y E M N E A
T_S K D K M_A E D E V E V Y I P Q F K L E E H Y E L R S I
L A K L G M T D V F C A A K A D L S G M N D G G L F L S T
V_F H_Q A_M V D V N E E G T E A A A A G T G V M T G R T G H
E E_H F T A D H P F L F F I R H N K T K S I L F L G R Y S
G G P Q F_V A D H P F L F L I M H K I T K_C I L F _ F G R F C
S P
S_P_

FTPA >fugu tPA candidate
(sequence length=524)
HTPA >gi|4505861| t-plasminogen activator [Homo sapiens]
(sequence length=524)

Number of matches = 294
Fraction of identities per length = 0.561069
FUTF >fugu tissue factor on scaf8956 238a  
(sequence length=264)

HSTF >gi|418689|pir||KFHU3 tissue factor precursor [validated] - huma  
(sequence length=295)

Number of matches = 88
Fraction of identities per length = 0.333333

MKCSALFLMLLLHCVRMETPAWPRVPFRPETAVARTLLGWVFQAQVA
FPLG >suspect fugu plasminogen from scaf9368 and scaf145
(sequence length=747)
HSPG  (16->810=795)>gi|4505881|ref|NP_000292.1| plasminogen [Homo sapiens
(sequence length=795)

Number of matches = 420
Fraction of identities per length = 0.562249

S V S G E V G Y T K T A G A W I L S L S G R K Y S V M T
S_G Q G E P L D D Y V N T Q G A S L F S V T K Q L G A G S
A V E C A R K C D A E T T F T C R \ S F I Y I E K D Q E C R T
I E E C A A K C E D E E F T C R / \ A F Q Y H S K E Q Q C_V I
V G Q N S K E I V L R R S S T A L Y E K K \ V Y L L E C V N
M A E N R K S S I I R M R D V V L F E K K / \ V Y L L E C K T
G I G T D Y R G T K S R T K T G K L C Q R W G V N S P H R P
N / Y K P Q T S P L A D L E S N F C R N P D A D S G G P W C Y
R / F S P A T H P S E G L E E N Y C R N P D N D P Q P W C Y_
T T D A N T R W E H C N V S A C S \ E D C I H C S G E D Y R G
T T D P E K R Y D Y C D I L E C E / \ E E C M H C S G E N Y D G_
K A S T T E K G Y T C Q R W D S K \ Y
K I S K T M S G L E C Q A W D S Q S P H A H G Y I P S K \ F P

L E G N Y C R N P D G D P R P W C F T T N P A K R W D
N K N L K Y N C R N P D R E L R P W C F T T D P N K R W E

F C D I P R C / T S E /
P P T I V P E T T C F T G E G G S Y R G
L C D I P R C / T T F
P P S S G P T Y Q C L K G T G E N Y R G

T I A V T E A G K T C Q S W S S Q T P H K H N R T P D N Y P
N V A V T V S G H T C Q H W S A Q T P H T H N R T P E N F P

C K /
C K / N L D E N Y C R N P D G K R A P W C H T T N S Q V R W E

Y C Q V P T C G D S A G P D D A V I P P E E E
Y C K I P S C D S S P V S T E Q L A P T A / P P E L T P V V Q

D C Y E D N G S S Y R G I T S Q T V S G K R C Q A W S S M T
D C Y H G D Q S Y R G T S T T T T G K C Q S W S S M T

P H S H L K T P K N F P N A / T S G G I F C R N P D N D R A P
P H R H Q K T P E N Y P N A / G L T M N Y C R N P D A D K G P

W C Y T T N P G R V W E Y C N L E K C S T N A P K T P G T
W C F T T D P S V R W E Y C N L K C S G T E A S V V

A T Q P Q S P T T E D S G Q T E R / D C K I G N E T Y R G K

I S I T I L G V T C Q A W S A Q S P H T H N S F T E E T H G
R A T T V T G T P C Q D W A A Q E P H R S I F T P E T N P

D K G L E S N / Y C R N P D G D V N G P W C Y T T D P N K K W
R A G L E K N / Y C R N P D G D V G G P W C Y T T N P R K L Y

D Y C D V P Q C A / A P S F D C G K P Q V E P K K C P G R V V

G G C V S K P H S W P W Q I S L R T S / R G I H F C G G T L I
G G C V A H P H S W P W Q V S L R T R / F G M H F C G G T L I

D R Q W V L T A A H C L E K / S T R P G Y Y K I V L G I H T E

R A I E A S K Q L R D L E K L V L G P N G A D I A L L K L Q
V N L E P H V Q E I E V S R L F L E P T R K D I A L L K L S

T / P A L L N E K V T P V C L P D K D Y I V P S H T E C Y V T
S / P A V I T D K V I P A C L P S P N Y V A D R T E C F I T

G W G E T Q / G T F G A G L L K E A Q L P V I E N K I C N R Y

A Y L N G R V K D H E M C A G N I E G G T D S C Q / G D S G G
E F L N G R V Q S T E L C A G H L A G G T D S C Q / G D S G G

PLVCNSNRFLQLQGVTSWGLGCANPMKPGV
PLVCFEDKLYILQGVTSWGLGCARPKNKPGV

YARVSKFIDWIKNTMELNYVRVSRFVTWIEGVMRNN

FUF7 >fugu factor7 candidate A
(sequence length=398)

HSF7 (44->466=423)>gi|182801|gb|AAA88040.1| coagulation factor VI
(sequence length=423)

Number of matches = 164
Fraction of identities per length = 0.412060

\/
\VFMKEANVFLHRTRRANFLFEELKAGNL
\VFPVTQEEAHGVLHRRRRANAFLEELRPGSL

DRDFKGKEFKSYEGAKEIFALPQQUESQGK
ERECKEEQCSEFAEEREIFKDAERTKLFWIS

AEQFIT//CKLSPCKNGATCTCRRFETYACKCA
YS//DGDQCASSPCQNGGSDKQLQSYICFC

NGFHGHNCDO//VRRTSNGCRYRNGGCEHFCR
PFEGERNCET//HKDDQLICVNENGGCEQYCS

EFPDRSYVCFCAPGYRLDKDNSTCLP//QVKV
DHTGTKRSCRUNCHGYSSLADGVSCTP//TVEY

PCGRLQILFSPRVINGLICFKGHPCKKIPILEKRNASKPGRIVGKVCVKGE

CPWO//AMLSENNITYTCAILSELSEQWVLTAAH
CPWO//VLLLLVNGAQLCGGTLPINTWVSAAAH

CVMRKPAHLFNVT//GEHDREIFKETEQHRCFKIKNWRNLIAVL//GEHDLSEHDGDEQSR

RVIKVLHPGYNKTSSDKDLAMLKHLRJPVK
RVAQVIIPSTYPGTTNHDIALLRLHQPV

LGLYVVPICLPAPA QNSTISRTLANRQSTVS
LTDHVVLCLP_ERTFSERTLAFVRFSLVVS

GWGRSLRSFPPATILOQRLTPLRVPQLQEC
GWGQLLDRGATALELMVLTNVPRLMTQDCLQ

RLHTKLNITRNMCLAGKLTGGRDACEG
QSRKVGDSPNTTEYMFCAAGYSDGSKSDKSCKG

DSGGPLVTYYKKTWFLTGTKVSVSWGGKGCANEDSGGPPHATHYRGTWYLTVSWSWGGCATV

LYGVYVRVTNFLDWIGNIIATNH
HFGVYTRVPSQIIEWLQKLRSERPRPGVLLR
A P F P

FUF9 >fugu factor 9 from scaf1343
(sequence length=473)
HUF9 >gi|22385321|gb|AAM96188.1| coagulation factor IX
(sequence length=461)

Number of matches = 215
Fraction of identities per length = 0.466377

M A R D F L L A L I A A L L E V S    GL P T E G S T G \ / V
M Q R V N M I A E S P G L I T I C L L G Y L L S A E C T \ / V_
F V S P Q A N M V L L R Q R R Y N S G H L E L Q K D N L
F L D H E N A N K I L N R P K R Y N S G K L E E F V Q G N L_
E R E C K E E Q C T M E E A R E V F E D D E K T \ /
E_R_E_C_M_E_E_K_C_S_F_E_E_A_R_E_V_F_E_N_T_E_R_T_\ / T E F W K Q
P D F S G R N C E I \ / E V S K Q C S V N N G G C S H F C V M Q
F G F E G K N C E L \ / D V T _ C N I K N G R C E Q F C K N S
G D I S V C H C A V G H R L G L D K S C E P T D \ / Q F S C
A D N K V C S C T E G Y R L A E N Q K S C E P A V \ / P F P C_
G H I N M S F S S K S N V Q R S L M Q K L E A N R T F S S
G R V S V S Q T S K_ L T R_ A E T V F P D
I L L G D Y S D N S T E L \ / D P Y W A F P T L P T I P E K E N
V D Y V N S T E A E T I L D N I T Q S T Q S F N D F
T D Q R I V G G D E A L P G E I P W Q \ V Q L N P P \ / A E P F C
T_ R_V V G G E D A K P G Q F P W Q \ / \ V V L N G K V D A F C_
G G S I V N E K W I V T A A H C V E T G V K I T V
V G \ / E H D V S K D E G P E R D H T V A E Q H I H F M Y D Y K
A G_ E H_/ N I E E T E H T E Q K R V I R I I P H H N Y N A A
I N K Y N H D I A L L L E D E P L V L N S Y V T P I C I A D
K D F T E T L L R E S T S S S L V S G W G R I K F F G L E T
K L Q K L E V P Y V D R T R C K Q S S R E Q V T R Y M F C A
V L Q Y L R V P L V D R A T C L R S T K F T I Y N N M F C A_
G Y Q L Q A K D S C Q G D S G G P H A T K Y K D T W F L T G
G F H E G G R D S C Q G D S G G P H V T E V G T S F L T G
I V S W G E E C A K D G K Y G I Y T R V S R Y Y P W I S Q K
T G L
T K L T

FUA2 >fugu uPA on scaf3932
(sequence length=454)
HUPA >human urokianase
(sequence length=431)

Number of matches = 175
Fraction of identities per length = 0.406032

MTFWRVLAAAFAILTTAEAVRTLLH/
MRALLARLLLCVLVVSDKGSNELHQ VPS/

CGLCACLHGGTSPVSLTSGEHMFCLCADGFQG
C_D C_L_N_G_T_C_V_S_N_K_Y_F_S_N_I_H_W_C_N_C_P_K_K_F_G_G_

KNCET/ VKNSEQCYEGVGLFLYRGTAQSQSESGR
QHECI/ DKSKTCEYEGNGHFRGKA_STDTMG_R_

TCRVPDPQT RESYLAS DINSGRH
PCLPWSATVLQQTYHAHRSDLQLGLGKH_

NYCR/ NLSRPPWCYVTKNQQLLWEYCAVP
NYCR/ N_PDNRRPPWCYVQVGLKPLVQECMVH

RCSSDL CRCIQPVPQKKNLDRVCDRGVLNLRL
DCADG/

CPHFAAPSPQAAPTPAHPAT/ SACGQRRRKRKQ
PSSIPPEELK FQCGQKTLPRLR

MKIVGTVATVESHPWVAAIFWRASKSEKV
FKII_GEGFTTIIENQPWFAAIYRRHRGGSVT

FRCCGSLLSSCWLVTAAHCF PDG/SHLKTRR
YVC_G_SL_SM_SP_C_W_V_I_SA_TH_CF/
ID YPKKED

FSVVLGKNALNERSKSTTEQKFGVEQIIVHR
YIVYLGRSRN QNTQGEMKEVENNLILHK

DFDNTDGNFNNIDI/ ALLKLKSTRGTCAKKSR
DYSADTLAHHNDI/ ALLKLKIRSKEGRCA_QPSR_

TVGSVCPLQHLLQP GLTCEIAGYGEQQ/H
TIQTICLP_SMYPDQF_GTSCEITTFGK픈NS/

GSWFRTRQYLRETQVNVISSDDVCRQEDYYGN
TDYLYPEQLKMTVVKLLISHRECQQPHYG_S

LITSNMFCAGQPDWSRDACK/ GDSGGPVLCE
EVTJKMLCACADPWKTDS_CQ/ GDSGGPLVCE_S
FUF7 >fugu f7 candidate B
(sequence length=428)

HSF7 (44->466=423)>gi|182801|gb|AAA88040.1| coagulation factor VI
(sequence length=423)

Number of matches = 187
Fraction of identities per length = 0.442080
FUF9 >fugu f9 candidate from scaf917
(sequence length=476)
HUF9 >gi|22385321|gb|AAM96188.1| coagulation factor IX
(sequence length=461)

Number of matches = 196
Fraction of identities per length = 0.425163

FUA1 >fugu uPA on scaf4367
(sequence length=392)
HUPA >human urokianase
(sequence length=431)

Number of matches = 164
Fraction of identities per length = 0.418367
FF7C >fugu f7 candidate C

(sequence length=426)

HSF7 (44->466=423)>gi|182801|gb|AAA88040.1| coagulation factor VI

(sequence length=423)

Number of matches = 173
Fraction of identities per length = 0.408983