J Mol Evol (1987) 26:358–369

Journal of Molecular Evolution

The Genetic Relationships between the Kringle Domains of Human Plasminogen, Prothrombin, Tissue Plasminogen Activator, Urokinase, and Coagulation Factor XII

Francis J. Castellino and John M. Beals

Department of Chemistry, University of Notre Dame, Notre Dame, Indiana 46556, USA

Summary. A computer-based statistical evaluation of the optimal alignments of the kringle domains of human plasminogen, human prothrombin, human tissue plasminogen activator, human urokinase, and human coagulation Factor XIIa, as well as the putative kringle of human haptoglobin, has been performed. A variety of different alignments has been examined and scores calculated in terms of the number of standard deviations (SD) of a given match from randomness. With the exception of human haptoglobin, it was found that very high alignment scores (8.9–23.0 SD from randomness) were obtained between each of the kringles, with the kringle 1 and kringle 5 regions of human plasminogen displaying the highest similarity, and the S kringle of human prothrombin and the human Factor XII kringle showing the least similarity. The relationships obtained were employed to construct an evolutionary tree for the kringles. The predicted alignments have also allowed nucleotide mutations in these regions to be evaluated more accurately. For those regions for which nucleotide sequences are known, we have employed the maximal alignments from the protein sequences to assess nucleotide sequence similarities. It was found that a range of approximately 40–55% of the nucleotide bases were placed at identical positions in the kringles, with the highest number found in the alignment of the two kringles of human tissue plasminogen activator and the lowest number in the alignment of the S kringle of prothrombin with the second kringle of tissue plasminogen activator. From both protein and nucleotide alignments, we conclude that haptoglobin is not statistically homologous to any other kringle.

Secondary structural comparisons of the kringle regions have been predicted by a combination of the Burgess and Chou–Fasman methods. In general, the kringles display a very high number of β-turns, and very low α-helical contents. From analysis of the predicted structures in relationship to the functional properties of these domains, it appears as though many of their functional differences can be related to possible conformational alterations resulting from amino acid substitutions in the kringles.

Key words: Kringle structures — Genetic relationships — Secondary structures — Alignments

Offprint requests to: F.J. Castellino

Introduction

The amino acid sequences of a number of proteins involved in both blood coagulation and fibrinolysis have been determined and their analyses have allowed much insight to be gained regarding the structure–function relationships of these proteins. One interesting feature of the primary structures of several proteins of this class was first revealed in bovine prothrombin upon recognition that a nonapeptide region was duplicated in the intact *prothrombin* molecule, along with a specific disulfide bridging pattern. As a result of alignment of these regions of the molecule, it was discovered that clear sequence homology existed between two peptide segments of approximately 80 amino acids in length, strongly suggestive of a partial gene duplication in this protein (Magnusson et al. 1975). These *structurally similar regions,* termed *kringles,* have since been pro-

posed to exist in several other related proteins. There appear to be two such areas of similarity in human prothrombin (Walz et al. 1977), five in human *plasminogen* (Sottrup-Jensen et al. 1978), two in human melanoma tissue *plasminogen activator* (Pennica et al. 1983), one each in human *urinary urokinase* (Verde et al. 1984) and human *Factor XIIa* (McMullen and Fujikawa 1985), and one proposed in human haptoglobin (Kurosky et al. 1980).

While visual observation is sufficient to reveal most of these sequence homologies, establishment of the complete genetic relationships between them requires a more exact manner of establishing their most favorable alignments. Direct comparisons have been made in several cases, with alignments based simply on the maximization of sequence identities. This process offers no systematic manner of foretelling where deletions may have occurred, and treats equally all identities and mismatches in the final alignment. With the knowledge that amino acids in related families of proteins differ with regard to their susceptibility to mutation (Dayhoff et al. 1978), and with the formulation of scoring matrices that allow some degree of quantitation of the significance of the entire range of possible amino acid substitutions (Dayhoff et al. 1978), we believed it important to examine in a rigorous fashion the nature of the alignments of the kringle regions of proteins that contained these unusual structures. A study such as this can contribute in a very meaningful manner to the ultimate understanding of the molecular evolution of these proteins and yield information on the retention of various structure–function relationships.

Experimental Procedures

Protein Fragments and Nucleic Acid Fragments. All amino acid sequences and mRNA sequences were taken from the protein and nucleic sequence database files of the Protein Identification Resource (PIR), except where noted otherwise. A tape containing this information was purchased from that source and installed on a VAX 11-750 computer. The proteins and regions employed were human plasminogen (HPg), residues 83–161 (PgK1), residues 165–242 (PgK2), residues 255–332 (PgK3), residues 357–434 (PgK4), and residues 461–540 (PgK5); human prothrombin (HF-II), residues 65–143 (IIKA) and residues 170–248 (IIKS); human haptoglobin (HuHP), residues 1–68 (HPF1); human melanoma tissue plasminogen activator (HTPA), residues 127–208 (TAK1) and residues 215–296 (TAK2); human urinary urokinase (HUK), residues 50–131 (UKK1); and the human Factor XIIa heavy chain, residues 198–276 (FXIIK1). This latter sequence was taken from the data of McMullen and Fujikawa (1985), and added to the database. The human plasminogen sequence present in this database was corrected for two differences between it and that determined from the cDNA sequence (Malinowski et al. 1984). The human haptoglobin sequence present in the database was corrected for four differences in residues 1–74 between it and the cDNA sequence of Raugei et al. (1983). With regard to the nucleic acid sequences of human urokinase and human plasminogen, not present in the data base, external files were created from the literature reports (Malinowski et al. 1984; Verde et al. 1984) and utilized in the same fashion as the database files. The published human plasminogen nucleic acid sequence does not include the region corresponding to kringles 1–3. Therefore, this segment was not included in nucleic acid alignments.

Alignment Procedures. In order to maximize the alignment of amino acid sequences from the different kringle regions, the computer program, ALIGN, purchased from the PIR, was employed after its installation on a VAX 11-750 computer. This program is based upon a published algorithm (Needleman and Wunsch 1970). In its operation, the real alignment score (R) is calculated from the mutation data matrix (Schwartz and Dayhoff 1979). This score is subtracted from the mean (M) of the highest scores for alignment of 100 random permutations of both sequences, and the result is divided by the standard deviation (SD) from this mean. At the discretion of the user, scoring penalties (NPEN) are imposed for sequence gaps. Additionally, a bias parameter (B) can be added to each term of the matrix to enhance or depress the real score. From the overall match score, in SD units, the probability of the true score being a chance occurrence can be found from standard error tables. An overall score >3.0 SD is supportive of a genetic relationship between the sequences under comparison.

Nucleic acid sequences were aligned according to the amino acid sequence gaps predicted above. For analysis of the resulting matches, a computer program was written in FORTRAN that allowed access to mRNA sequence data files and insertion of codon gaps at the desired positions. This was compiled on a VAX 11-750 computer. A final count of the number of codons in which specified numbers of base changes had occurred, and the number of identical nucleic acids in the final alignment was produced as the output file.

Secondary Structure Calculations. Secondary structures for the above fragments were predicted from the rules of Chou and Fasman (1978) using a program written in PASCAL and compiled on a VAX 11-750 (Castellino et al. 1986). Our Chou–Fasman program includes subroutines for preliminary determinations of secondary structure, followed by boundary analyses, resolution of β-turn overlaps with other secondary structures, resolution of helix-sheet overlaps, and determination of shortened β-sheet regions, prior to a final prediction. A second algorithm (Burgess et al. 1974) was also employed for these predictions with aid of a program which we developed in FORTRAN, and compiled on a VAX 11-750 computer.

Results

Protein Alignments

A computer-based statistical method has been utilized to provide assistance in decisions as to the optimal manner of alignment of the homologous kringle regions of a series of fibrinolytic and blood coagulation proteins. In employing this procedure, it is necessary to include scoring penalties for gaps placed in the sequences under comparison, otherwise an unreasonable number of breaks will be included in an attempt to maximize the number of positions of sequence identities in the final result. Also, alignment scores of some distantly related sequences sometimes benefit from addition of a constant to all terms of the scoring matrix. Therefore,

Table 1. Alignment scores for kringle structures in various proteins[a]

	PgK1[b]	PgK2[c]	PgK3[d]	PgK4[e]	PgK5[f]	IIKA[g]	IIKS[h]	TAK1[i]	TAK2[j]	UKK1[k]	FXIIK1[l]
PgK1	31.8	—	—	—	—	—	—	—	—	—	—
PgK2	17.8	37.6	—	—	—	—	—	—	—	—	—
PgK3	17.9	19.2	35.5	—	—	—	—	—	—	—	—
PgK4	18.9	20.5	21.4	34.1	—	—	—	—	—	—	—
PgK5	23.0	18.1	17.6	19.0	31.9	—	—	—	—	—	—
IIKA	20.1	17.6	16.8	16.3	18.3	35.0	—	—	—	—	—
IIKS	15.8	12.4	15.3	14.4	12.8	14.8	34.6	—	—	—	—
TAK1	12.6	14.8	13.6	13.9	9.4	11.7	9.4	31.9	—	—	—
TAK2	13.6	14.0	14.5	14.4	13.9	9.0	10.2	20.1	32.5	—	—
UKK1	10.9	13.5	12.5	13.0	10.5	9.7	10.4	17.0	16.6	34.6	—
FXIIK1	13.5	14.2	13.7	16.7	13.5	10.1	8.9	15.1	19.0	15.7	32.0

[a] Real score (R) − mean score from 100 random runs (M)/SD of M; where B = 2 and NPEN = 15
[b] HPg kringle 1, protein residues 83–161
[c] HPg kringle 2, protein residues 165–242
[d] HPg kringle 3, protein residues 255–332
[e] HPg kringle 4, protein residues 357–434
[f] HPg kringle 5, protein residues 461–540
[g] HFII kringle A, protein residues 65–143
[h] HFII kringle S, protein residues 170–248
[i] HTPA kringle 1, protein residues 127–208
[j] HTPA kringle 2, protein residues 215–296
[k] HUK kringle 1, protein residues 50–131
[l] HFXII kringle 1, protein residues 198–276

in the utilization of this technique we have chosen a series of progressively increasing scoring matrix bias parameters (B) and break penalty parameters (NPEN), in order to determine their effects on the results obtained. In many cases, the nature of the alignments was influenced by the choice of these parameters, and several different matches were found to be possible, each with a very high score.

Table 1 summarizes the optimal alignment scores for all combinations of kringles, using a value for B, of 2.0, and a very high value for NPEN, of 15.0. Here, the lowest overall match score obtained was 8.9 SD from randomness, for the alignment of the kringle S region of prothrombin (IIKS) with the kringle region of human Factor XIIa (FXIIK1). This highly significant value shows that the structures are indeed genetically related. On the other hand, the highest score obtained under these matrix bias conditions was 23.0 SD, for the alignment between the kringle 1 region of human plasminogen (PgK1) and the kringle 5 (PgK5) domain of this same protein.

While these data indicate that PgK1 is most similar to PgK5, further analysis of the results suggests that the kringle 2 region of human plasminogen (PgK2) and plasminogen kringle 3 (PgK3) can be best aligned with the kringle 4 segment (PgK4) of this protein. Considering the prothrombin kringles, KA (IIKA) is most similar to PgK1 with a score of 20.1 SD and prothrombin kringle KS (IIKS) exhibits virtual equal number of matches relative to PgK1 and IIKA, with alignment scores of 15.8 SD and 14.8 SD, respectively. The first kringle of tissue plasminogen activator (TAK1) is most highly similar to the kringle 2 region of the same molecule (TAK2), alignment score of 20.1 SD. In addition, TAK2 aligns best with FXIIK1, alignment score of 19.0 SD. The single kringle region of urokinase (UKK1) shows nearly equal similarity with TAK1 and TAK2, with scores of 17.0 SD and 16.6 SD, respectively.

Alignments of the kringle domains, referred to TAK2, are summarized in Fig. 1.

We have performed analyses with a wide range of values of B and NPEN (data not shown). In the most extreme case, when B was set at 0 and NPEN at 6, or when B was set to 20 and NPEN at 20, all alignment scores ranged between 7.9 SD and 19.3 SD, values which still indicate that a large departure from randomness exists in the calculated alignments. However, in the cases that differed from those in Table 1, either too many gaps required insertion, or residue identities were minimized. The former problem reduced the number of critical cysteine residue matches, and these cases were not believed to reflect proper alignments. Therefore, we have employed the alignments based on a bias parameter of 2 and break penalty of 15 to provide, as closely as possible, the actual gap insertion and residue mutations that have been directed by evolutionary processes. The other alignments, although not reported here, were used to support the evolutionary relationships between the kringles.

In Table 2, the number of gaps and gap insertion points in the kringle alignments have been listed. Although alignment scores provide a numerical point total based on the number of identities and gaps necessary to align two proteins, the actual gap insertion points provide the detailed information necessary for evaluation of the evolutionary relationships between the kringles. In analyzing the data, three subclasses of kringles can be assembled in which no gaps occur in their optimal alignments. The underlying premise in the assembly of our evolutionary relationships between the kringles is that

```
TAK2   CYFGNGSAYR'GTHSLTESGA'SCLPWNSMI 'LIGKVYTAQN'PSAQALGLGK'HNYCRNPDGD'A KPWCHVLK'NRRLTWEYCD'VPSC
PgK1   CKTGDGKNYR'GTMSKTKNGI'TCQKWSSTS-'PHRPRFSPAT'HPSEG   LE'ENYCRNPDND'PQGPWCYITD'PE KRYDYCD'ILEC
PgK4   CYHGDGQSYR'GTSSTTTGK'KCQSWSSMT-'PHRHQKTPEN'YPNAG   LT'MNYCRNPDAD'K-GPWCFTTD'PS VRWEYCN'LKKC
PgK5   CMFGNGKGYR'GKRATTVTGT'PCQDWAAQEP'HRHSIFTPET'NPRAG   LE'KNYCRNPDGD'VGGPWCYTTN'PR KLYDYCD'VPQC
IIKA   CAEGLGTNYR'GNVSITRSGI'ECQLWRSRY-'PHKPEINSTT'HPGAD   LQ'ENFCRNPDSS'ITGPWCYTTD'PT ARRQECS'TPVC
FXIIK1 CYDGRGLSYR'GLARTTLSGA'PCQPWASEA-'TYR   NVTA'EQARNWGLGG'HAFCRNPDND'I-RPWCFVLN'RDRLSWEYCD'LAQC
PgK2   CMHCSGENYD'GKISKTMSGL'ECQAWDSQS-'PHAHGYIPSK'FPNKN   LK'KNYCRNPDRE'L-RPWCFTTD'PN KRWELCD'IPRC
PgK3   CKLGTGENYR'GNVAVTVSGH'TCQHWSAQT-'PHTHNRTPEN'FPCKN   LD'ENYCRNPDGK'R-APWCHTTN'SQ VRWEYCK'IPSC
IIKS   CVPDRGQQYQ'GRLAVTTHGL'PCLAWASAQ-'AKALSKHQDFN'SAVQ   LV'ENFCRNPDGD'EEGVWCYVAG'KP GDFGYCD'LNYC
TAK1   CYEDQGISYR'GTWSTAESGA'ECTNWNSSA-'LAQKPYSGRR'PDAIRLGLGN'HNYCRNPDRD'S-KPWCYVFK'AGKYSSEFCS'TPAC
UKK1   CYEGNGHFYR'GKASTDTMGR'PCLPWNSAT-'VLQQTYHAHR'SDALQLGLGK'HNYCRNPDNR'R-RPWCYVQV'GLKPLVQECM'VHDC
```

Fig. 1. Primary structure alignments for the kringles (K) of human plasminogen (Pg), human prothrombin (II), human tissue plasminogen activator (TA), human urokinase (UK), and human Factor XII (FXII). The alignments are based on TAK2, with gaps represented as spaces. In alignments not requiring gaps at positions 30 or 62 in TAK2, a dash (–) was inserted to indicate a no gap site and that the peptide segment was continuous in sequence. The ' is used to indicate every 10th position, including gaps.

Table 2. Sequence break assignments for kringle structures in various proteins[a]

Sequence 1	Breaks in sequence 2										
	PgK1[b]	PgK2[c]	PgK3[d]	PgK4[e]	PgK5[f]	IIKA[g]	IIKS[h]	TAK1[i]	TAK2[j]	UKK1[k]	FXIIK1[l]
PgK1	—	59	57	57	—	—	—	62	62	38; 39; 64	59
PgK2	—	—	—	—	—	—	—	—	—	—	—
PgK3	—	—	—	—	—	—	—	—	—	—	—
PgK4	—	—	—	—	—	—	—	—	—	—	—
PgK5	35	35; 60	36; 60	36; 60	—	36	45–48	60	30; 61	62	34; 60
IIKA	—	59	57	57	—	—	—	62	62	62	59
IIKS	—	55	58	58	32–34	—	—	63; 69	63	63	60
TAK1	46–48; 73	46–48; 72	34–36; 67	35–37; 72	43; 50; 75	46–48; 77	47–49; 76–77	—	—	—	30–33
TAK2	47–49; 73	46–48; 72	47–49; 67	48–50; 67	48–50; 75	46–48; 68	30; 31; 49; 68	—	—	—	39–41
UKK1	47–51; 77	46–48; 74	47–49; 74	48–50; 74	43; 50; 68	46–48; 68	47–49; 69	—	—	—	33–35
FXIIK1	70	69	68	68	72	70	70	45	—	—	—

[a] In the convention used, the breaks extend the sequences at the point of insertion and are included in the final numbering of the amino acid residues. B = 2 and NPEN = 15

[b] HPg kringle 1, protein residues 83–161
[c] HPg kringle 2, protein residues 165–242
[d] HPg kringle 3, protein residues 255–332
[e] HPg kringle 4, protein residues 357–434
[f] HPg kringle 5, protein residues 461–540
[g] HFII kringle A, protein residues 65–143
[h] HFII kringle S, protein residues 170–248
[i] HTPA kringle 1, protein residues 127–208
[j] HTPA kringle 2, protein residues 215–296
[k] HUK kringle 1, protein residues 50–131
[l] HFXII kringle 1, protein residues 198–276

the addition or deletion of genetic material is a less favorable route, as compared to point mutations in the gene, to a functionally mutated form of the kringle containing approximately the same number of residues. Therefore, alignments with the fewest number of gaps have been assumed to be the most likely paths of evolution. With this rationale, the kringles appear to be divisible into a number of subgroups of very close homologies. Subclass 1 is composed of PgK1, IIKA, and IIKS. Subclass 2 consists of PgK2, PgK3, and PgK4. The last subclass, 3, is composed of TAK1, TAK2, and UKK1. Kringle PgK5 and FXIIK1 require gaps for maximum alignment with each of the other kringles. Suggested relationships for these last two kringles, as well as the interrelationships between the three subclasses, based upon gap insertion points in addition to the alignment scores and the number of identities retained after maximal alignment, are shown in Fig 2, and will be discussed below.

Table 3 presents a listing of the number of residues conserved after break insertions in the optimal alignments of the kringle domains. As expected, the kringles that exhibited the best alignment scores and

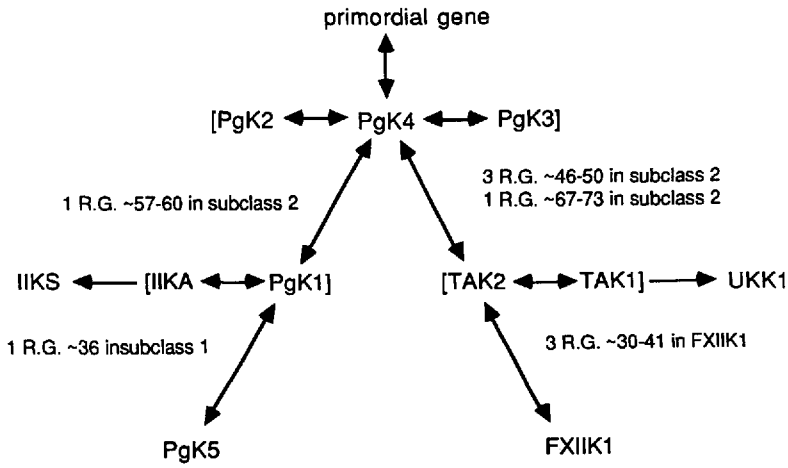

Fig. 2. The evolutionary relationship between kringles. Kringles located within brackets represent relationships that could be determined. The notation #R.G refers to the number of residues per gap followed by the approximate region for gap insertion. Subclasses are as referred to in the text.

Table 3. Number of residues conserved for kringle structures in various proteins[a]

	PgK1[b]	PgK2[c]	PgK3[d]	PgK4[e]	PgK5[f]	IIKA[g]	IIKS[h]	TAK1[i]	TAK2[j]	UKK1[k]	FXIIK1[l]
PgK1	79	—	—	—	—	—	—	—	—	—	—
PgK2	39	78	—	—	—	—	—	—	—	—	—
PgK3	37	41	78	—	—	—	—	—	—	—	—
PgK4	43	39	41	78	—	—	—	—	—	—	—
PgK5	45	37	41	42	80	—	—	—	—	—	—
IIKA	41	35	35	35	35	79	—	—	—	—	—
IIKS	26	25	24	27	30	24	79	—	—	—	—
TAK1	28	29	25	29	27	30	26	82	—	—	—
TAK2	29	30	34	34	34	26	28	44	82	—	—
UKK1	29	27	24	28	30	27	29	39	39	82	—
FXIIK1	27	26	26	32	31	25	27	35	41	33	79

[a] Number of identical matches after breaks have been included. B = 2 and NPEN = 15
[b] HPg kringle 1, protein residues 83–161
[c] HPg kringle 2, protein residues 165–242
[d] HPg kringle 3, protein residues 255–332
[e] HPg kringle 4, protein residues 357–434
[f] HPg kringle 5, protein residues 461–540
[g] HFII kringle A, protein residues 65–143
[h] HFII kringle S, protein residues 170–248
[i] HTPA kringle 1, protein residues 127–208
[j] HTPA kringle 2, protein residues 215–296
[k] HUK kringle 1, protein residues 50–131
[l] HFXII kringle 1, protein residues 198–276

a pattern with no sequence breaks, in general, yield the highest number of identical residues. PgK1 has 41 residues that are homologous with IIKA. The other kringle in the first subclass of kringles, IIKS, shows only 26 identities with PgK1 and 24 identities with IIKA, an observation that suggests a distinct evolutionary divergence of these domains. The subclass of kringles PgK2, PgK3, and PgK4 shows striking similarities, with 50% or greater identity between them. The last subclass, containing TAK1, TAK2, and UKK1, has greater than 48% identity, with TAK1 and TAK2 having 44 of 82 identical residues in their optimal alignments. In addition, TAK2 shows the greatest similarity with FXIIK1, having 51% of their residues conserved. Other notable related sequences include PgK1 and PgK5, containing 45 identities, PgK1 and PgK4, with 41 identities, and PgK3 and PgK4, each having 34 identities with TAK2, after maximal alignment. It should be noted that these latter similarities are all between different subclasses of kringles and provide a means by which to relate the kringles.

It has been proposed that a region of human haptoglobin, comprising residues 1–68 of its α-chain, is homologous to the fifth kringle region of human plasminogen (Kurosky et al. 1980). We have evaluated this proposal by the above methods, employing a variety of values of B and NPEN, and the results are presented in Table 4. It is clear from the data that this region of haptoglobin possesses similarities no greater than expected by chance for all of the kringles, with the exception of PgK3, where an alignment can be constructed that is statistically greater than a random genetic occurrence, but even in this case the similarity is very limited.

Secondary Structure Predictions

The two predictive schemes (Burgess et al. 1974; Chou and Fasman 1978) were used in combination

Table 4. Statistical alignments of human haptoglobin with various kringle structures

Sequence		SD from randomness with the indicated bias parameters		
First	Second	B = 2, NPEN = 15	B = 15, NPEN = 15	B = 15, NPEN = 5
HPF1[a]	PgK1[b]	−0.97	−0.02	0.83
HPF1	PgK2[c]	1.48	0.74	0.71
HPF1	PgK3[d]	3.31	2.77	2.45
HPF1	PgK4[e]	1.18	0.09	0.70
HPF1	PgK5[f]	0.82	1.08	0.54
HPF1	IIKA[g]	−0.99	−0.63	−0.02
HPF1	IIKS[h]	−1.13	0.27	0.39
HPF1	TAK1[i]	0.00	0.39	0.29
HFP1	TAK2[j]	1.20	1.25	1.43
HPF1	UKK1[k]	0.52	0.41	1.07
HPF1	FXIIK1[l]	0.28	0.41	1.09

[a] HuHP, protein residues 1–68
[b] HPg kringle 1, protein residues 83–161
[c] HPg kringle 2, protein residues 165–242
[d] HPg kringle 3, protein residues 255–332
[e] HPg kringle 4, protein residues 357–434
[f] HPg kringle 5, protein residues 461–540
[g] HFII kringle A, protein residues 65–143
[h] HFII kringle S, protein residues 170–248
[i] HTPA kringle 1, protein residues 127–208
[j] HTPA kringle 2, protein residues 215–296
[k] HUK kringle 1, protein residues 50–131
[l] HFXII kringle 1, protein residues 198–276

to predict the secondary structures of the proteins of interest. The predictions for the kringle regions of the proteins that are presented were excised from information on the entire protein. The Chou–Fasman approach was used to calculate all possible helical, sheet, and turn regions using preliminary determinations and boundary analysis subroutines. The Burgess method produced final predictions for helix and sheet structures, as well as all possible β-turns. In addition, a graphic printout of individual residues existing in helices (hi), β-sheets (bi), and coils (ci) was generated.

Ordered conformational regions that were predicted identically by both algorithms were automatically executed. In addition, comparative β-turn conformations that were displaced by ± 1 residue were also considered. Regions not predicted by the Burgess algorithm were analyzed with the graphic printout. Those areas with bi and hi values greater than the cutoff value for such structures were considered as possible conformations and correlated with those predicted from the Chou–Fasman algorithm. Assignments that revealed both β-sheet and α-helical potential were based primarily upon overlap resolution methods of Chou and Fasman (1978). Normalized β-turn potentials were calculated for those conformations that were predicted by both methods. This potential was calculated by summing turn values, ci and p_t, that have been normalized using both techniques, employing maximal turn potentials of 0.145 for the Burgess algorithm and 1.55×10^{-3} for the Chou–Fasman algorithm. In cases in which overlapping turns existed, identical turns with the highest normalized potential were executed in the prediction. If there were no identical turns, displaced turns with the highest normalized potential were assigned.

Secondary structure alignments of the kringle domains, referred to TAK2, are shown in Fig. 3.

Discussion

Protein Alignments

Table 1 summarizes the optimal alignment scores for all combinations of kringles, using a value for B, of 2.0, and a high value for NPEN, of 15.0. In no case did the alignments affect identities between Cys residues. This approach tended to yield maximal alignments containing a minimal number of sequence breaks. The break locations presented in Fig. 1 represent the approximate gap sites determined by averaging the gap locations that we assigned between all kringle regions and the TAK2 domain. This latter region was selected as the reference structure for structural reasons presented below. In certain cases, gap sites at positions 30 and 62 were necessary in TAK2 for maximal alignments to be achieved. Our rationale in the choice of these particular B and NPEN parameters was that the regions examined were obtained from proteins from the same mammalian species, were derived from proteins with logical functional relationships, and were of approximately the same number of amino acid residues. Therefore, we concluded that many gaps in sequence were not likely to be actual genetic events, and we placed a large differential between B and NPEN in order to minimize the number of breaks inserted in the most probable statistical alignments.

Manual qualitative attempts have been made in the past in order to align some of these kringle regions, primarily based upon maximizing the number of sequence identities. In one example, an alignment of PgK5 and UKK1 has been proposed that involved sequence gaps at positions 19, 27, 51, and 68 in PgK5, and at positions 13, and 61 in UKK1 (Gunzler et al. 1982). Employing a B of 2.0 and an NPEN of 15, we calculate the match score for this alignment to be 7.7 SD, as compared to our score of 10.5 SD. Therefore, while the manual attempt has provided a credible result, it is not the most statistically probable alignment, and that of Table

```
TAK2   RRRTTTTRTT'TTBBBBRRRR'RRRTTTBBB 'BBRRRRRRRR'RRRBBBBBTT'TTRTTTTTTT'T TTTBBBBT'TTBBBBBBRR'RTTT
PgK1   RRRTTTTRTT'TTRRRTTTTR'RBBBRTTTT-'RRTTTTTTTT'RRRRR   RR'RTTTTRRRRT'TTTRBBBBTT'TR RBBBBBR'RRRR
PgK4   RRRTTTTRTT'TTRRBBBBRR'RTTTTRRRT-'TTTRRRRTTT'TRTTT   BB'BBTTTRRRRR'R-TTTBBBBT'TT BBBBBBB'BRRR
PgK5   RRRTTTTRTT'TTRRBBBBBR'RTTTTRRRRR'RRRRRRTTTT'RRRRR   RR'RTTTTRTTTT'RRTTTBBBTT'TB BBBBBBR'BBBB
IIKA   RRRRRRRRTT'TTBBBBBTTB'BBBBBTTTT-'RRRRRRTTTT'TTTTR   RR'RRRRRRTTT'TRTTTBBBTT'TR RRRRRRR'RRRR
FXIIK1 TTTTTTRTTT'TBBBBBBRRT'TTTRRRRRR-'BBB   BBRH'HHHHRRRRRR'RRRRRTTTTR'R-TTTBBBBT'TTTRRRRBBB'BBBB
PgK2   RRRRRRRRTT'TTRRRRRRRR'RRRRTTTTT-'TTTRRRTTTT'RRTTT   TR'RRRTTTTRRR'B-BBBBBRRT'TT TRRRRRR'BBBB
PgK3   BBRTTTTRTT'TTBBBBBRRR'RRRRRRRRT-'TTTRRRTTTT'RRTTT   TR'RTTTTRTTTT'R-TTTTRRRB'BB BRRRRRR'RRRR
IIKS   BBRTTTBBBBRRBBBBBRRRR'RRHHHHHHHH-'HHHHHHRTTT'TBBBB   BR'RRTTTTTTTT'R-RBBBBBRR'RR RTTTTRR'RRRR
TAK1   BBRTTTBBBB'BBBRRRRRRR'RRTTTTRRR-'RRRRRTTTTR'RRRBBBBBRT'TTTRRRRTTT'T-TTTBBBBB'RTTTTRRRRR'RRRR
UKK1   RRTTTTRRRR'RRRTTTTRRR'RBBBBBRRB-'BBBBRRRRTT'TTRBBBBBRR'RTTTTRTTTT'B-BBBBBBBB'RRRRBBBBBB'BRRT
```

Fig. 3. Secondary structures of the kringle domains of various proteins, as predicted from Chou–Fasman calculations. The structures represented are: α-helices (A), β-sheets (B), coils (C), and β-turns (T). All kringles have been aligned with TAK2 as the reference. In cases where the comparison does not require a sequence gap in TAK2 at residue sites 30 and 62, a (–) has been inserted to indicate peptide sequence continuation. The ' is used to indicate every 10th position, including gaps. PgK1 through PgK5 represents kringles 1 through 5, respectively, in human plasma plasminogen. IIKA and IIKS refer to kringles A and S, respectively, in human plasma prothrombin. TAK1 and TAK2 are kringles 1 and 2, respectively, in human melanoma tissue plasminogen activator. UKK1 is the single kringle of human urinary urokinase. XIIK1 represents the lone kringle in human Factor XIIa.

1 possesses a higher probability of reflecting the true genetic events that have occurred. Another such manual attempt has been forwarded (Magnusson 1979). Here, the sequence similarities of the two bovine prothrombin kringles and the five human plasminogen kringles were compared. In the case of the alignments of PgK1 and PgK2 with PgK5, the proposed gap (Magnusson 1979) at position 36 of PgK1 coincides with the statistically determined gap generated in this report at position 35 of PgK1. Further, in the comparison of PgK2 with PgK4, and PgK3 with PgK4, we find that a more favorable alignment in both cases would involve no sequence gaps, rather than the two placed in each of the pairs under consideration by the manual alignment (Magnusson 1979). Other such refinements are possible in the alignments proposed by simple visualization procedures.

An effort to quantitate alignment scores of some of these kringle structures has been made (Patthy et al. 1984). Here, a system was devised that assigned all amino acid sequence identities a score of +10, except for Cys, which was scored as +20. Sequence gaps were provided scores of −25. A computer system was apparently not employed in order to maximize scores for all possible alignments within a given comparison, and all possible comparisons beween kringle structures were not made. Most of our statistically predicted optimal alignments differ from those previously reported (Patthy et al. 1984). As one example, the alignment of PgK1 and TAK1 can be considered. Here, a proposed alignment (Patthy et al. 1984) that contains gaps at positions 24, 46–48, 52, 65, 72, and 73 in PgK1 and at positions 23, 60, 61, and 74 in TAK1 has been proposed. Our alignment, proposed in Table 2, contains fewer sequence gaps. Further, employing a B value of 2 and an NPEN of 15, the match score for the alignment proposed by these investigators is 3.9 SD from randomness. Our value, of 12.6 SD, under these conditions, is much more favorable. Since genetic trees of the relatedness of the kringles have been constructed (Patthy et al. 1984) based upon these less-than-optimal alignments, they must be viewed with some skepticism.

Nucleic Acid Alignments

One very useful application of these more soundly constructed alignments, with genetically probable gaps properly inserted, is that the nucleotide base changes that occurred in these similar structural regions can be accurately assessed and commented upon. Table 5 provides a summary of this information for the particular kringle regions that have experimentally determined nucleotide sequences, i.e., PgK4 and PgK5 (Malinowski et al. 1984), IIKA and IIKS (Degen et al. 1983), TAK1 and TAK2 (Pennica et al. 1983), UKK1 (Verde et al. 1984), and HPF1 (Raugei et al. 1983). In the majority of the cases, the nucleic acid alignments presented in Table 5 augment the intersubclass relationships that have been implied by the protein sequence alignments and do not indicate that a change in any of the relationships suggested above is necessary. For example, it is seen from Table 5 that the plasminogen activator kringles possess well-conserved nucleotide sequences in the final alignments. In com-

Table 5. Nucleotide base differences in mRNA in the alignments of the kringle regions of various proteins

Sequence		No. of codons requiring changes of[a]				No. of identical bases (total)
First	Second	0	1 base	2 bases	3 bases	
PgK4[b]	PgK5[c]	18	26	23	11	129 (234)
PgK4	IIKA[d]	15	31	17	15	124 (234)
PgK4	IIKS[e]	14	23	21	20	109 (234)
PgK4	TAK1[f]	17	15	27	19	108 (234)
PgK4	TAK2[g]	20	21	20	17	122 (234)
PgK4	UKK1[h]	15	21	27	15	114 (234)
PgK5	IIKA	17	23	19	20	116 (237)
PgK5	IIKS	16	23	21	19	115 (237)
PgK5	TAK1	15	21	25	18	112 (237)
PgK5	TAK2	21	21	24	13	129 (237)
PgK5	UKK1	17	20	21	21	112 (237)
IIKA	IIKS	16	14	28	21	104 (237)
IIKA	TAK1	15	21	21	21	108 (234)
IIKA	TAK2	9	27	25	17	106 (234)
IIKA	UKK1	15	19	27	17	110 (234)
IIKS	TAK1	13	16	27	21	98 (231)
IIKS	TAK2	14	20	23	21	105 (234)
IIKS	UKK1	17	15	19	27	100 (234)
TAK1	TAK2	26	20	18	18	136 (246)
TAK1	UKK1	24	17	21	20	127 (246)
TAK2	UKK1	22	23	19	18	121 (246)
HPF1[i]	PgK5	9	11	26	15	75 (261)

[a] Codons matching sequence gaps have been ignored in the count
[b] HPg kringle 4, protein residues 357–434
[c] HPg kringle 5, protein residues 461–540
[d] HFII kringle A, protein residues 65–143
[e] HFII kringle S, protein residues 170–248
[f] HTPA kringle 1, protein residues 127–208
[g] HTPA kringle 2, protein residues 215–296
[h] HUK kringle 1, protein residues 50–131
[i] HuHP, protein residues 1–68

paring maximal alignments of TAK1, TAK2, and UKK1, 22–26 of the codons are preserved intact, and 17–23 codons only require a single base change. Overall 121–136 of the total of 246 nucleotides are identical. These strong similarities argue for the placement of these domains in the same subclass in the evolutionary tree proposed. Similarly, the weaker similarity between TAK1 and PgK5, observed from protein alignments (Table 1), is also reflected in the nucleic acid alignments. From the data of Table 5, it is seen that in this case 43 of the 79 codons require either two or three base changes in the final alignment. These two kringles are placed in different subclasses in Fig. 2. On the other hand, a larger number of nucleotide matches is seen with PgK5 and TAK2 than was noted in amino acid sequences. Overall, when different proteins are considered, the weakest similarities are observed with the prothrombin kringles and those from the two plasminogen activators. The strongest similarities are present between the two types of plasminogen activators.

The maximal alignment between HPF1 and PgK5 obtained from protein sequences was also employed to evaluate nucleotide matches between these two regions. As seen in Table 5, only nine of the 261 codons were identical in the alignments. A total of 75 identical bases were observed, suggesting that the similarity between HPF1 and PgK5 was very weak. This argues further that HPF1 is most likely not a kringle structure.

Evolutionary Relationships

These statistically most probable alignments have also been used herein to assess evolutionary relationships (Fig. 2) between the kringles, thus allowing us to speculate on possible structure–function correlations in the less studied kringles. Our evolutionary tree was generated based upon three considerations: (1) the relationships are due to a divergent evolutionary process involving gene duplication as evidenced by the high degree of similarity with respect to amino acid residues, (2) the deletion and/or insertion of genetic material proceeds by a minimal manipulative route, thus the relationships with the fewest gaps are favored, and (3) the number of residue identities is indicative of relatedness. Initially, the kringles can be separated into three subclasses: (1) PgK1, IIKA, IIKS; (2) PgK2, PgK3, PgK4; and (3) TAK1, TAK2, UKK1. Separation of the various kringles into these subclasses is based upon the data suggesting that no gaps are necessary for their optimal alignments. By this criterion, kringles PgK5 and FXIIK1 cannot be included in any one subclass since both require breaks to align with any of the other kringles.

The statistical alignment of kringles in subclass 1 indicates that PgK1 and IIKA are the most similar. Yet, it is not obvious if IIKS is closer on an evolutionary time scale to PgK1 or IIKA since alignments yield similar results with 26 and 24 identities and alignment scores of 15.8 SD and 14.8 SD, respectively. These relationships are then presented schematically as [PgK1 ≒ IIKA] → IIKS.

The second subclass mentioned, containing PgK2, PgK3, and PgK4, yields some of the highest alignment scores. PgK2 aligns most favorably with PgK4 and PgK3. It is noted that alignment of PgK2 with PgK3, as opposed to PgK2 and PgK4, yields two more identities, yet has a lower alignment score, of 19.2 SD. In each case a *no gap* alignment has been used for the comparison. Thus, the former comparison yields matches in residues that are less likely to mutate, as suggested by the mutation database (Dayhoff et al. 1978). The latter comparison provides more sequence identities than the former, but

the mutations are in less probable directions, thus lowering the alignment score. These relationships are thus represented as [PgK2 ≒ PgK4 ≒ PgK3].

The final subclass of kringles, TAK1, TAK2, and UKK1, shows a similar relationship as in subclass 1 where TAK1 and TAK2 represent the most favorable alignment. However, it is unclear how UKK1 is related to these other two kringles, with each alignment yielding 39 identities and scores of ~17.0 SD. The schematic representation of these relationships are, therefore, represented as [TAK1 ≒ TAK2] → UKK1.

Intersubclass relationships were based primarily on minimizing the number of gaps necessary to optimally align each kringle subclass, with direct relationships based on scores and identities conserved. The assembled relationships are represented in Fig. 1. PgK1 and IIKA are probably closer on an evolutionary time scale to subclass 2 since the alignment requires only a one-residue break in subclass 2, in the vicinity of residue 57 (Table 2), for optimal alignment. The alignment of PgK1 or IIKA with subclass 3 requires a minimum of three breaks. The most direct relationship between subclass 1 and subclass 2 resides in PgK1 and PgK4 where 43 identities are seen (Table 3), with an alignment score of 18.9 SD (Table 1). Although subclass 2 shows evolutionary homology to subclass 1, it also provides a minimal manipulative route to subclass 3, requiring only two gaps. The first gap is a three-residue break in subclass 2 occurring within a region from residues 46 to 50 (Table 2). The second gap involves a single residue in subclass 2, probably occurring within a region encompassing residues 67–77 (Table 2). The most direct relationship is observed between PgK4 or PgK3 with TAK2, having 34 identities each and alignment scores of ~14.5 SD. This proposed relationship is supported by the fact that maximal alignment of subclass 1 with subclass 3 requires a similar gap insertion pattern as subclass 2, with a three-residue gap in subclass 1 within a region ranging from residues 46 to 50 and a one-residue gap in subclass 1 within a region ranging from residues 67 to 77, in addition to another one-residue gap in subclass 3 between residues 62 and 64. Thus, it appears that subclass 2 bridges the other two subclasses.

A possible criticism of this alignment procedure is that exact gap sites are not conserved. The fact that more specific sites cannot be located is probably due in part to a divergent evolution mechanism that involves point mutations which result in amino acid changes that influence the statistical alignments. Thus, slight variation in gap insertion points are seen with the ALIGN program. Despite this problem, the location of break regions are reasonably apparent.

The two remaining kringles PgK5 and FXIIK1 appear to align optimally with PgK1 and TAK2, respectively. The relationship of PgK5 to PgK1, of subclass 1, is supported by the observation that the alignment of subclass 2 requires two single-residue gaps that correspond to the gap necessary for alignment to subclass 1, between residues 57 and 60 in subclass 2, and a gap for the alignment of subclass 1 to PgK5, at approximately residue 36 in subclass 2.

FXIIK1 is most closely aligned with TAK2, with 41 identities and only one three-residue gap occurring in FXIIK1 within a region ranging from residues 30 to 41. This same gap is inserted regardless of whether the kringle domain of subclass 3 is aligned with FXIIK1.

The final evolutionary relationships (Fig. 2) are consistent with a known functional property of some of the kringles. Fibrin-binding capabilities have been proposed for PgK1 (Lerch et al. 1980), PgK4 (Lerch et al. 1980; Cole and Castellino 1984), TAK2 (Ichinose et al. 1986), and to a lesser extent for PgK5 (Sottrup-Jensen et al. 1978), supporting the relationships generated by the alignments between the proposed subclasses. In addition, it would be expected that IIKA, and possibly FXIIK1, would show the same binding capability, but preliminary results (Ichinose et al. 1986) indicate that this is not the case.

The evolutionary tree, presented in Fig. 2, differs somewhat from that previously suggested (Patthy et al. 1984) in that we are able to establish distinct evolutionary relationships between the kringles. The evolutionary relationships reported in this communication suggest that all of the kringles have been derived from a single gene, but the occurrence of kringles in other proteins has arisen by the duplication of the genetic material responsible for PgK4. In addition, it appears from the alignment scores calculated that in proteins with multiple kringles, gene duplication of an initial existing kringle has occurred, e.g., TAK2 was duplicated yielding TAK1. It has been suggested that all the kringles are derived directly from a singular primordial gene (Patthy et al. 1984). While this is certainly a possible explanation, it appears to be an inefficient route to kringles that possess the functions of fibrin binding and ϵ-amino caproic acid (EACA) binding. Our scheme suggests an evolution of a functional protein domain that has been incorporated over time into other proteins providing a property beneficial for survival. It is noted that a number of the relationships reported (Patthy et al. 1984) agree with those reported here, particularly the close relationships between the plasminogen kringles and IIKA as well as the homology between the kringles in the plasminogen activators.

Secondary Structure Predictions

Figure 3 illustrates our predictions of the secondary structures of these homologous kringle regions, according to the rules of Chou–Fasman (Chou and Fasman 1978) and Burgess et al. (1974). While any reference structures could have been chosen, in this example the alignments have all been referred to TAK2, which contains a known fibrin-binding site (Ichinose et al. 1986), with the following amino acid gap locations: (1) sites 30 and 62 in TAK2; (2) regions 46–48 and 72 in the plasminogen and prothrombin kringles; and (3) region 34–36 in FXIIK1. In all the alignments, the gaps have been counted as occupying a position. In cases in which the aligned kringle did not require gap insertions at either residue 30 or 62 in TAK2, a dash was inserted to indicate the absence of a gap site and that the peptide sequence was continuous. The first four kringle structures have been purported to contain lysine-, EACA-, and/or fibrin-binding sites and are grouped at the top of the figure to facilitate inter-kringle comparison of secondary structure.

It can be readily seen that while genetic homology has been retained, the amino acid substitutions that have occurred have potentially important consequences, since several areas of conformational dissimilarities exist. For example, it is believed that the kringle 1 domain of human plasminogen and kringle 2 of human tissue plasminogen activator are responsible for the interaction of these proteins with fibrinogen and fibrin (Thorsen et al. 1972; Wiman and Collen 1978; Rijken and Collen 1981; Wallen et al. 1982; Ichinose et al. 1986). This allows plasminogen activation to occur favorably on the fibrin surface with this particular activator (Thorsen 1975). However, despite the occurrence of a homologous kringle in human urokinase and human Factor XIIa, these proteins do not bind in a functionally significant fashion to fibrinogen and fibrin. Thus, the evolutionary differences that exist in these kringle areas have allowed changes, most likely of the conformational kind, that significantly alter the functional properties of these molecules.

In another example, human plasminogen is known to bind the effector molecule, EACA, most probably via its kringle regions. However, plasminogen kringle 1 interacts with this molecule very tightly (Lerch et al. 1980), kringle 4 binds EACA with lesser affinity (Lerch et al. 1980; Cole and Castellino 1984), and kringle 5 does so much more weakly than either kringle 1 or kringle 4 (Sottrup-Jensen et al. 1978). Urokinase, prothrombin, and Factor XIIa are not known to specifically interact with EACA, at least not in a fashion that has a functional consequence.

With these insights, and with the assumption that the lysine-, EACA-, and fibrin-binding sites are the same, it is now possible to align the kringles PgK1, PgK4, PgK5, and TAK2 and attempt to discover homologous regions in the secondary structure that may be important to these interactions. Two homologous regions can be located with respect to secondary structure: (1) a region including residues 1–12, containing two β-turns, and (2) a region stretching from approximately 65 to 77 that contains a β-sheet/β-turn/β-sheet sequence of conformations. It is interesting to note that native plasminogen residues, Asp_{412} and Arg_{425} (positions 60 and 75, including gaps, in the K4 sequence), reported to be involved in the lysine-binding site of PgK4 (Trexler et al. 1982), are located in this latter region of homology. It is also interesting to note that IIKA, IIKS, and FXIIK1 do not bind to fibrin (Ichinose et al. 1986). In Fig. 3, it is shown that these latter kringles, with the possible exception of FXIIK1, lack secondary structural homology in the region between residues 66 and 77. This suggests that the secondary structure in this region is important to binding of the above effector molecules. The secondary structures of FXIIK1 and PgK2 both exhibit a possible β-sheet/β-turn/β-sheet sequence with the second β-sheet region starting closer to the COOH-terminal end of the domain. With regard to the first 12 positions of these kringles, only PgK3 retains the first two β-turns seen in the EACA-binding kringles.

While showing high turn potential, we observe that a nonapeptide region, containing residues 52–60, although homologous at the primary sequence level, shows little evidence for homology at the secondary structure level. However, another region of secondary structure similarity, consisting of residues 66–77, shows strikingly less primary sequence similarity. The amino- and carboxyl-terminal ends of this peptide display the greatest similarity between kringles and the β-turn located in this region shows much greater sequence diversity. Since the amino acid substitutions that have occurred in this area have allowed retention of the β-sheet/β-turn/β-sheet conformation in kringles that bind the effector molecules listed above, further support is obtained for the importance of that particular secondary structural arrangement to the binding process. If this region is indeed essential, FXIIK1, which also contains that same predicted secondary structure in this region, should also bind these ligands. While it is enticing to speculate that these two regions of secondary structure are essential for the fibrin- or lysine-binding site, the inability of other kringles to participate in this binding could readily be explained by conformational differences in the relevant site that are induced by point mutations in other parts of the molecule. It should be noted that in addition to the lack of secondary structure similarity in the region of amino acids 66–77 between the kringles

that do not bind the effector molecules, primary structure similarities reveal that critical substitutions for the Asp_{60} residue in PgK4 have also been made in kringles IIKA, IIKS, and UKK1. The Arg_{75} residue in PgK4 is conserved to some extent in each of the kringles, with the exception of IIKS, with Lys found as the common positively charged substitute, occurring within two positions of site 75.

An example of this latter situation is noted when comparing UKK1 and FXIIK1. The large degree of amino acid sequence similarity between these regions is not reflected to the same degree in their secondary structures, which have noticeable differences. This is readily explainable in a manner typified by examination of the structure prediction for residues 8–11 in each sequence. Here, the sequence, Ser-Tyr-Arg-Gly, in FXIIK1, is predicted as a β-turn. In UKK1, the same residues are Phe-Tyr-Arg-Gly, a sequence with a low β-turn potential. Thus, the substitution of a Phe for a Ser residue in position 1 alters the structural prediction considerably. Other examples of such effects with conservative substitutions are possible. This type of situation most likely has allowed some functional diversity to be present in genetically similar regions of these molecules. This again points to important conformational differences in regions of molecules with extremely strong genetic relationships, that are brought about by amino acid substitutions.

While results from secondary structural predictions have been very useful for attempting to explain functional differences between the kringles, it is obvious that consideration of tertiary structural differences will also be important to a more detailed analysis of structure–function relationships of these highly interesting domains. However, this awaits detailed x-ray structural analysis of all of the kringles. To date, only one such structure exists, that for IIKA (Park and Tulinsky 1986), a kringle without reported interesting functional properties.

The comparison of the secondary structure predicted for IIKA is in reasonably good agreement with that determined from the crystal structure of this area (Park and Tulinsky 1986). These investigators have shown that the two intradisulfide bonds in this kringle form a sulfur cluster and that the peptide segments surrounding these disulfide regions exist as antiparallel β-sheets. Two strands, Asn_{52}–Asp_{58} and Gly_{63}–Thr_{68}, correspond to the two regions of highest residue conservation. The first β-sheet is not predicted by our approach since this segment contains residues of low β-sheet potential. This indicates that interactions with the other strand may induce a β-sheet in this peptide segment. The Gly_{63}–Thr_{68} sheet corresponds to the first strand predicted between positions 66 and 77 in our study. As noted by Park and Tulinsky (1986), these two β-strands are separated by a β-turn, as we predict. The remainder of this kringle structure has been inferred by Park and Tulinsky (1986) to exist in β-sheet and β-turn conformations. The inability to produce exact assignments has been limited by the contorted folding in the kringle structure. The predictions in this work support the evidence of Park and Tulinsky (1986) that the remainder of the structure exists as β-sheets and β-turns, with the exception of FXIIK1 and IIKS, which are predicted to contain short regions of α-helix.

Alignment with Haptoglobin

The protein and nucleic acid alignments generated in this study suggest little-or-no homology sequence similarity between any of the kringles and HPF1, an observation that differs from a report (Kurosky et al. 1980), which claimed that the most numerous matches were between HPF1 and PgK5. In an attempt to improve alignment scores, we extended the comparisons of the various kringles with the sequence of the human haptoglobin α-chain, consisting of residues 1–74, which terminates at the last Cys position in this chain. The results obtained suggested that no substantial matches exist.

In conclusion, we have produced a quantitative evaluation of the genetic relationships between the functionally significant kringle domains of human plasma proteins that are involved in blood coagulation and fibrinolysis. In all cases, alignment scores are extremely high, indicating that these segments are indeed related by a process other than chance. The manner of alignment that produced the highest match was not difficult to decide in most cases, and, in those that were ambiguous, intuitive reasoning allowed a justifiable decision to be reached. The overall process of producing match scores has allowed a comparison to be made regarding the similarities of these different proteins, a feature that could not be attained in a rigorous fashion from past attempts centered on maximizing sequence identities. It is most reasonable to propose that a primordial gene that coded for a kringle loop has been replicated many times, altered, and recombined to produce part of the structures of plasminogen, prothrombin, tissue plasminogen activator, urokinase, and Factor XIIa. In general, the kringles that are most similar reside within the same proteins, and the plasminogen kringles are most similar to those of prothrombin. The single kringle domain in urokinase shows the least similarity with that same region in other proteins, except for the tissue plasminogen activator kringles and the single kringle structure in Factor XIIa, with which it is most highly similar, and with which it can be believed to be homologous.

Acknowledgments. These studies were supported by grants HL-13423 and HL-19982 from the National Institutes of Health. We express our gratitude to Professor Kenneth Grant and Mr. Joseph Weber for their assistance with computer programming.

References

Burgess AW, Ponnuswamy PK, Scheraga HA (1974) Analysis of conformations of amino acid residues and prediction of backbone topography in proteins. Isr J Chem 12:239–286

Castellino FJ, Sator de Serrano V, Powell JR, Johnson WR, Beals JM (1986) Examination of the secondary structure of the kringle 4 domain of human plasminogen. Arch Biochem Biophys 247:312–320

Chou PY, Fasman GD (1978) Predictions of secondary structures of proteins from their amino acid sequence. Adv Enzymol 47:45–148

Cole KR, Castellino FJ (1984) The binding of antifibrinolytic amino acids to kringle 4-containing fragments of plasminogen. Arch Biochem Biophys 229:568–575

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 2. National Biomedical Research Foundation, Washington DC, pp 345–352

Degen SJF, MacGillivray RTA, Davie EW (1983) Characterization of the complementary deoxyribonucleic acid and gene coding for human prothrombin. Biochemistry 22:2087–2097

Gunzler WA, Steffens GJ, Otting F, Kim SA, Frankus E, Flohe L (1982) The primary structure of the high molecular mass urokinase from human urine. The complete amino acid sequence of the A-chain. Hoppe-Seyler's Z Physiol Chem 363: 1155–1165

Ichinose A, Takio K, Fujikawa K (1986) Localization of the binding site of tissue-type plasminogen activator to fibrin. J Clin Invest 78:163–169

Kurosky A, Barnett DR, Lee T-h, Touchstone B, Hay RE, Arnott MS, Bowman BH, Fitch WM (1980) Covalent structure of human haptoglobin: a serine protease homolog. Proc Natl Acad Sci USA 77:3388–3392

Lerch PG, Rickli EE, Lergier W, Gillessen D (1980) Localization of individual lysine binding regions in human plasminogen and investigations on their complex forming properties. Eur J Biochem 107:7–13

Magnusson S (1979) Thrombin and prothrombin. In: Blomback B, Hansen LA (eds) Plasma proteins. John Wiley & Sons, New York, pp 254–276

Magnusson S, Petersen TE, Sottrup-Jensen L, Claeys H (1975) Complete primary structure of prothrombin: isolation and reactivity of ten carboxylated glutamic residues and regulation of prothrombin activation by thrombin. In: Reich E, Rifkin DB, Shaw E (eds) Proteases and biological control. Cold Spring Harbor Press, New York, pp 123–149

Malinowski DP, Sadler JE, Davie EW (1984) Characterization of a complementary deoxyribonucleic acid coding for human and bovine plasminogen. Biochemistry 23:4243–4250

McMullen BA, Fujikawa K (1985) Amino acid sequence of the heavy chain of human alpha factor XIIa (activated Hageman Factor). J Biol Chem 260:5328–5334

Needleman SB, Wunsch CD (1970) A general method applicable to search for similarities in amino acid sequence of 2 proteins. J Mol Biol 48:443–453

Park CH, Tulinsky A (1986) Three dimensional structure of the kringle sequence: structure of prothrombin fragment 1. Biochemistry 25:3977–3982

Patthy L, Trexler M, Banyai L, Varadi A (1984) Kringles: modules specialized for protein binding. FEBS Lett 171:131–136

Pennica D, Holmes WE, Kohr WJ, Harkins RN, Vehar GA, Ward CA, Bennett WF, Yelverton E, Seeburg PH, Heyneker HL, Goeddel DV, Collen D (1983) Cloning and expression of human tissue type plasminogen activator cDNA in *Escherichia coli.* Nature 301:214–221

Raugei G, Bensi G, Colantuoni V, Romano V, Santoro C, Costanzo F, Cortese R (1983) Sequence of human haptoglobin cDNA: evidence that the a and b subunits are coded by the same mRNA. Nucleic Acids Res 11:5811–1819

Rijken DC, Collen D (1981) Purification and characterization of the plasminogen activator secreted by human melanoma cells in culture. J Biol Chem 256:7035–7041

Schwartz RM, Dayhoff MO (1979) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington DC, pp 353–358

Sottrup-Jensen L, Claeys H, Petersen TE, Magnusson S (1978) The primary structure of human plasminogen: isolation and characterization of two lysine-binding fragments and one 'mini'-plasminogen by elastase-catalyzed-specific limited proteolysis. In: Davidson JF, Rowan RM, Samama MM, Desnoyers PC (eds) Fibrinolysis and thrombolysis, vol 3. Raven Press, New York, pp 191–209

Thorsen S (1975) Differences in binding to fibrin of native plasminogen and plasminogen modified by proteolytic degradation. Influence of omega-carboxylic acids. Biochim Biophys Acta 393:55–65

Thorsen S, Glas-Greenwalt P, Astrup T (1972) Differences in binding to fibrin of urokinase and tissue plasminogen activator. Thromb Diath Haemorrh 28:65–74

Trexler M, Vali Z, Patthy L (1982) Structure of the e-amino caproic acid binding sites of human plasminogen. J Biol Chem 257:7401–7406

Verde P, Stoppelli MP, Galeffi P, DiNocera P, Blasi F (1984) Identification and primary sequence of an unspliced human urokinase poly(A)+ RNA. Proc Natl Acad Sci USA 81:4727–4731

Wallen P, Bergsdorf N, Ranby M (1982) Purification and identification of 2 structural variants of porcine tissue plasminogen activator by affinity adsorption on fibrin. Biochim Biophys Acta 719:318–328

Walz DA, Hewett-Emmett D, Seegers WH (1977) Amino acid sequence of human prothrombin fragment-1 and fragment-2. Proc Natl Acad Sci USA 74:1969–1972

Wiman B, Collen D (1978) Molecular mechanism of physiological fibrinolysis. Nature 272:549–550

Received April 1, 1987/Revised May 13, 1987