View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Volume 157, number 2 FEBS 0523 July 1983

Adaptation of plasminogen activator sequences to known protease structures

W. Straßburger, A. Wollmer, J.E. Pitts[+], I.D. Glover[+], I.J. Tickle[+], T.L. Blundell[+], G.J. Steffens°, W.A. Günzler°, F. Ötting° and L. Flohé°

Lehr- und Forschungsgebiet Struktur und Funktion der Proteine, Abteilung Physiologische Chemie, RWTH Aachen, Forckenbeckstraße, D-5100 Aachen, FRG, [+]Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, England and *°Grünenthal GmbH, Forschungszentrum Aachen-Eilendorf, Abteilung Proteinchemie, Zieglerstr. 66, D-5100 Aachen, FRG*

Received 4 May 1983

The sequences of urokinase (UK) and tissue-type plasminogen activator (TPA) were aligned with those of chymotrypsin, trypsin, and elastase according to their 'structurally conserved regions'. In spite of its trypsin-like specificity UK was model-built on the basis of the chymotrypsin structure because of a corresponding disulfide pattern. The extra disulfide bond falls to cysteines 50 and 111d. Insertions can easily be accommodated at the surface. As they occur similarly in both, UK and TPA, a role in plasminogen recognition may be possible. Of the functional positions known to be involved in substrate or inhibitor binding, Asp 97, Lys 143 and Arg 217 (Leu in TPA) may contribute to plasminogen activating specificity. PTI binding may in part be impaired by structural differences at the edge of the binding pocket.

Urokinase *Tissue-type plasminogen activator* *Chymotrypsin* *Trypsin* *Model building* *Computer graphics* *Sequence alignment* *β-Structure formation*

1. INTRODUCTION

Urokinase is a fibrinolytic serine protease known in two active forms, high and low molecular mass urokinase, which are identical with respect to their B chain, where the catalytic residues are located. Their mass difference is due to a cleavage in the A chain leaving 21 out of originally 157 amino acids as chain A1 which remains linked to the B chain by a single disulfide bridge [1]. It is apparent from the UK sequence as well as from the cleavage of the physiological substrate behind an Arg residue that UK has a trypsin-like specificity. However, it still remains to be revealed by which factors specificity is narrowed to the activating cleavage of plasminogen. Since the sequence of UK is known [2–4] the primary structure of another protein sharing that specificity was elucidated; i.e., tissue-type plasminogen activator [5]. This enables the two sequences to be compared with one another and with those of homologous serine proteases. By calculation of the inherent secondary structure potential and model building on the basis of an X-ray determined homologous structure it is attempted to 'extrapolate' the sequence information into three dimensions. The obvious aim of the study is to possibly recognize striking features which may have a bearing on specificity or on the reported failure of UK to be inhibited by PTI [6], and also

Abbreviations: UK, urokinase; HUK, high molecular mass urokinase; LUK, low molecular mass urokinase; TPA, tissue plasminogen activating enzyme; CT, chymotrypsin; T, trypsin; PTI, pancreatic trypsin inhibitor; E, elastase; SCR, structurally conserved region

Nomenclature: Three- and one-letter symbols are used for amino acids [7]

Published by Elsevier Science Publishers B.V.
00145793/83/$ 3.00 © 1983 Federation of European Biochemical Societies

to raise more specific questions to be answered by experiment.

2. MATERIALS AND METHODS

The amino acid sequences of HUK and LUK have recently been determined [1–4] and those of TPA [5], CT, T, and E were taken from the literature [7]. The potential of β-structure formation was calculated with the parameters of Chou and Fasman averaging over four adjacent residues [8].

Model building was carried out using an interactive computer graphics system, an Evans and Sutherland Picture System II with an extensively modified version [9] of the FRODO program [10]. This program allows the replacement of residues, bond rotations for manual elimination of bad contacts, and regularization for defined parts of a structure [11] with continuous updating of coordinates.

The atomic coordinates of γ-CT were taken from the Brookhaven Protein Data Bank [12], those of T and the T/PTI complex were kindly provided by Dr W. Bode [13,14].

3. RESULTS AND DISCUSSION

The alignment of the sequences of LUK and of the protease moiety of TPA with one another and with those of other mammalian serine proteases is straightforward (see table 1). UK as well as TPA exhibit the characteristic pattern of structurally conserved regions [16]; however, they both have a 4-residue insertion in SCR 102–115. There are no substitutions incompatible with the distribution of hydrophobic core or solvent accessible regions in the chains of CT and T. If in core-segments a cer-

Table 1

Alignment of the sequences of human tissue-type plasminogen activator, human urokinase, bovine chymotrypsin, bovine trypsin, and porcine elastase using the one-letter symbols [7] and the chymotrypsinogen numbering [15] (insertions are indicated a,b,c...)

Protein	Residues (block 1)	Residues (block 2)	Residues (block 3)	Residues (block 4)
Tissue-type plasminogen activator	CGLRQYSQPQFR IKGGLFADIASHPWQAAIFAKHRRSPGERFLCGGILISSCWILSAAHCFQERFPPHHL	TVILGRTYRVVPGEEEQKFEVEKYIVHKEFDDDTY DNDIALLQLKSDSSRCAQESSVVRTVCLPPADLQLPD	WTECELSGYGKHEALSPFYSERLKEAHVRLYPSSRCTSQHLLNRTVTDNMLCAGDTRSGGPQANLHDACQGDSG	GPLVCLNDGRMTLVGIISWGLG CGQKDVPGVYTKVTNYLDWIRDNMRP
Urokinase	CGQKTL RPRFK IIGGEFTTIENQPWFAAIYRRHR GGSVTYVCGGSLISPCWVISATHCFIDYPKKEDY	IVYLGRSRLNSNTQGEMKFEVENLILHKDYSADTLAHHNDIALLKIRSKEGRCAQPSRTIQTICLPSMYNDPQF	GTSCEITGFGKENSTDYLYPEQLKMTVVKLISHRECQQPHYYGSEVTTKMLCAAD PQ WKTDSCQGDSG	GPLVCSLQGRMTLTEIVSWGRG CALKDKPGVYTRVSHFLPWIRSHTKEENGLAL
Chymotrypsin	CGVPAI QPVL IVNGEEAVPGSWPWQVSLQDKT GFHFCGGSLINENWVVTAAHCGVT TSD	VVVAGEFDQGSSSEKIQKIKIAKVFKNSKYNSLTI NNDITLLKLSTA ASFSQTVSAVCLPSASDDFAA	GTTCVTTGWGLTRY ANTPDRLQQASLPLLSNTNCKK YWGTKIKDAMICAGA SGVSSCMGDSG	GPLVCKKNGAWTLVGIVSWGSS TCSTS TPGVYARVTALVNWVQQTLAAN
Trypsin	IVGGYTCGANTVPYQVSLNS GYHFCGGSLINSQWVVSAAHCYKS GI	QVRLGQDNINVVEGNQQFISASKSIVHPSYNSNTL NNDIMLIKLKSA ASLNSRVASISLPT SCASA	GTQCLISGWGNTKSSGTSYPDVIKCLKAPILSNSSCKS AYPGQITSNMFCAGY LQ GGKDSCQGDSG	GPVVCS GKLQGIVSWGSG CAQKNKPGVYTKVCNYVSWIKQTIASN
Elastase	VVGGTEAQRNSWPSQISLQYR S GSSWAHTCGGTLIRQNWVMTAAHCVDRE LTF	RVVVGEHNLNQNNGTEQYVGVQKIVVHPYWNTDDVAAGYDIALLRLAQS VTLNSYVQLGVLPRAGTIIAN	NSPCYITGWGLTRTN GQLAQTLQQAYLPTVDYAICSSSSYWGSTVKNSMVCAGG DG VRSGCQGDSG	GPLHCLYNGQYAVHGVTSFVSRLGCNVTRKPTVFTRVSAYISWINNVIASN

Boxes designate structurally conserved regions [16]: (●) catalytic residues; (*i*) inhibitor (PTI) binding residues; (disulfide-bond arrow symbols) disulfide bonds; (▼) site of carbohydrate attachment (in UK)

tain position is substituted differently for UK and TPA the chemical nature of the substituent residues is found to be closely similar. The largest differences between the two plasminogen proteases are a 6-residue insertion in TPA behind position 185 and a 6-residue extension at the C-terminus of UK. It seems noteworthy that Asn 145 which carries the carbohydrate chain of UK is Glu in TPA, a residue unknown as a site of carbohydrate attachment.

The disulfide pattern of UK and TPA corresponds to that of CT, however, there is one additional cysteine pair, 50–111d. In spite of the T-like specificity of UK it was the disulfide pattern which led us to model UK based on the structure of CT.

The obvious interest of the study lies in the recognition of those residues which are responsible for the plasminogen-activating specificity and for the failure to bind PTI. These should be residues common to the sequences of both UK and TPA but different from their counterparts in T and CT, respectively. For residues involved in substrate or inhibitor binding as known from the X-ray determined structure of the complexes with T and CT [14,17], major differences in their chemical nature are encountered in positions 39, 41, 97, 143 and 217. The two Asp in 97 and the two Lys in 143 may be of special importance.

The predominant type of main-chain conformation realized in the serine protease family is the β-structure which is organized into two 'barrels'. Therefore, we calculated the potential of β-structure formation for UK, TPA and CT.

Fig.1 shows the appropriately normalized absolute differences in β-potential for UK and TPA *vs* CT. The plausibility of the resulting picture confirms the adequacy of the sequence alignment: The large differences (amongst them of course those due to insertions) are found almost exclusively outside the SCRs and are always found for both, UK and TPA. This is part of the general observation that the distribution of the differences in β-potential between UK or TPA, respectively, and CT are nearly symmetrical. The largest differences within the SCRs are found between positions 16 and 22, 135 and 145, and at the C-terminus. The very small differences between positions 180 and 232 are due to a particularly high density of invariant residues in this stretch of the three sequences.

In model building most replacements were feasible without necessitating changes in the main-chain conformation of the underlying γ-CT structure. This was not the case for a few glycines with positive ϕ and ψ angles; e.g., Gly$\longrightarrow$Phe 59 in close proximity to the catalytic His 57 also occurring in TPA. Problems were also to be expected for the accommodation of new prolines as, for instance, Phe$\longrightarrow$Pro 114 which is further complicated by the adjacent 50–111d disulfide bond and the inserted 111a–d loop.

Replacements by Pro in UK, by the way, do not

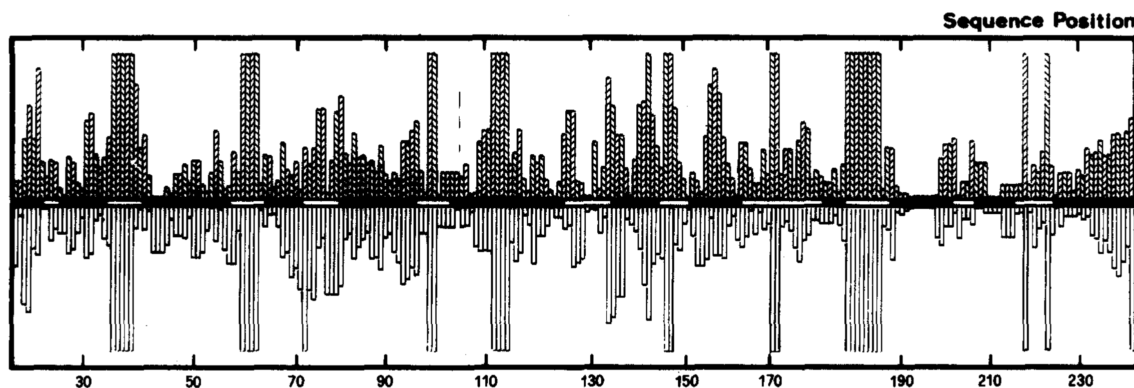

Fig.1. Comparison of urokinase and tissue-type plasminogen activator *vs* chymotrypsin with respect to their potential of β-structure formation [8]. Absolute normalized potential differences UK minus CT are plotted upwards, TPA-minus-CT downwards. (Insertions are given maximum values. The non-linear appearance of the sequence-position scale is due to the use of the chymotrypsinogen numbering.) Black sections on the abscissa are indicating structurally conserved regions.

have a correspondence in TPA and vice versa. Insertions are facilitated by the fact that, including the one in SCR 102–115, they can be accommodated at the surface. The protrusion into the solvent means that for these segments the model coordinates within certain limits are arbitrary. If the disulfide pattern of CT is accepted for UK (and TPA) the 50–111d position of the extra disulfide bond is automatic. Its relatively smooth formation in model building seems to confirm the pattern. Several of the interactions, polar as well as apolar, resulting from modelling UK based on CT seem to be favourable for structure stabilization such as the new salt bridges Glu⟶Arg 70/Ile⟶Glu 80, Asp⟶Arg 72/Asp⟶Glu 153, Lys⟶Glu 84/Ser⟶Arg 109, and Ala⟶Glu 86/Lys = Lys 107 or the extended hydrophobic cluster Gln⟶Phe 30/Gln⟶Tyr 34/His⟶Tyr 40/Gln⟶Leu 73/Trp⟶Phe 141.

Recently, the interaction of T and PTI was taken as a model for the recognition between prothrombin and its activating enzyme, blood clotting factor X_A [18]. The specificity was attributed to two Glu residues at the third position before and behind the cleaved peptide bond of prothrombin selected by salt bridge formation with Arg 143 and Lys 62 in factor X_A. Complementary relationships of this kind between corresponding positions in plasminogen and UK which would explain the restriction of the T-like specificity to plasminogen activation do not exist. Positions 41, 99, 143, 190, 217, 219 and 221, which count amongst the sites in contact with inhibitors or substrates, in UK and TPA are occupied by amino acids which are encountered also in blood clotting proteases [19]. The combination of amino acids in those positions, however, is different from one protease to the other and their pattern may well be important for specificity. The only substitutions which are unique are those in position 217, Arg in UK and Leu in TPA. Thus, these side chains at the top entrance of the binding pocket should, if at all, interfere by their size rather than by the positive charge which only occurs in UK.

From the fact that both, UK and TPA, are very similar with respect to the localization of insertions, one could infer that these might have a role in plasminogen recognition. However, it should be kept in mind that model building leaves these surface appendices with conformational alternatives. Another appendix which, though near the binding pocket, seems irrelevant for specificity is the carbohydrate chain since it is lacking in TPA [5].

The PTI binding area of T and CT prior to their exact description by the X-ray structure of the

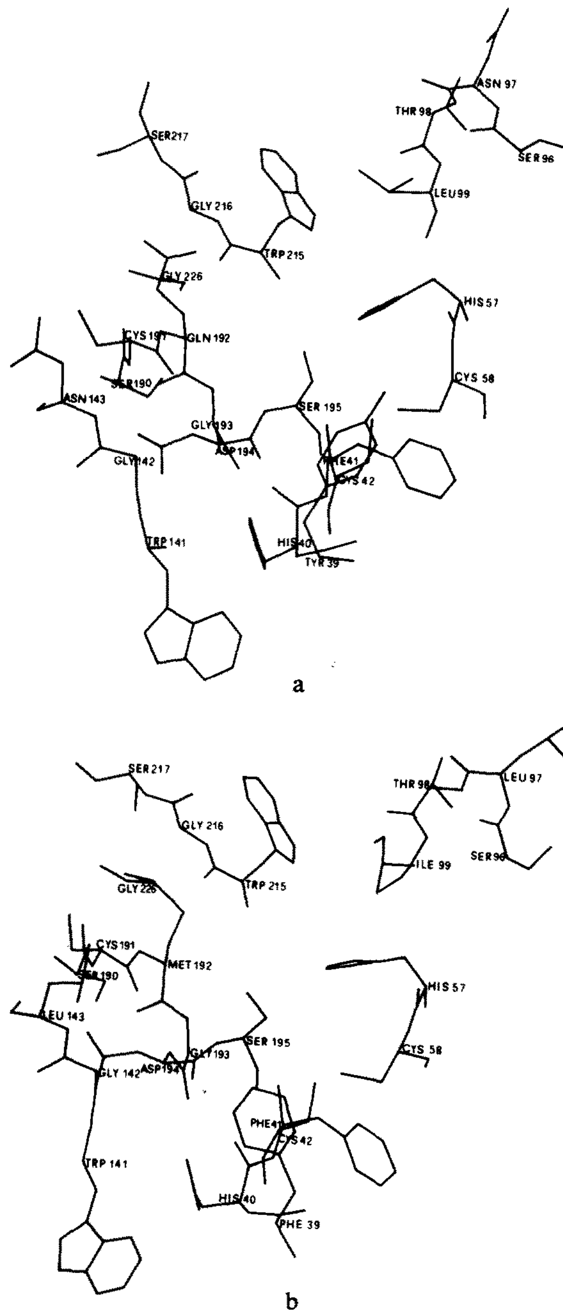

Fig.2. Residues of trypsin (a) and chymotrypsin (b) involved in the binding of pancreatic trypsin inhibitor.

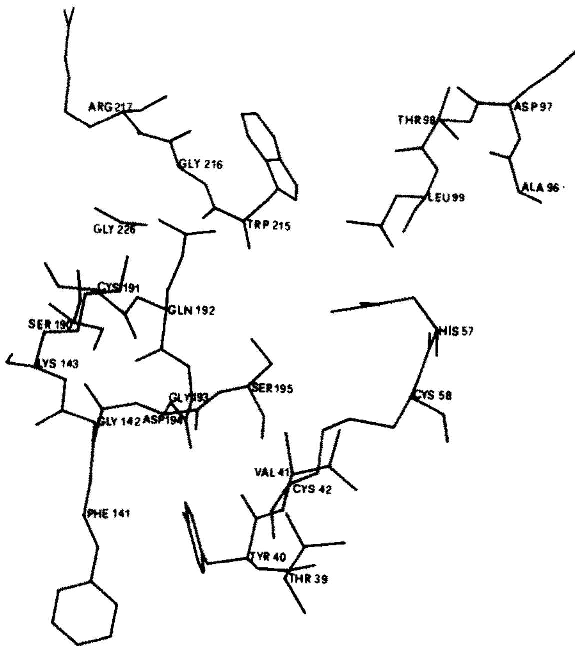

Fig.3. Residues of urokinase corresponding to those which in trypsin and chymotrypsin bind pancreatic trypsin inhibitor.

T/PTI complex [14] was also explored by model building studies [20]. Only a small number of residues was shown to participate in the numerous van der Waals' contacts. The arrangement of these residues in T, CT and UK juxtaposed in fig.2a,b and 3 clearly indicates that replacements do not occur in the closer surroundings of the active site. Amongst the places where they do occur the edge of the binding pocket around residue 40 is the only one displaying a similarity between T and CT on the one hand as opposed to UK on the other. However, a striking steric feature, e.g., a single side chain responsible for the reported absence of inhibition by PTI and, hence, for the obstruction of PTI binding, cannot be identified. The non-binding may be due to residues even further remote from the active site (i.e., outside fig.3) or to a multitude of minor geometric inadequacies.

Although we are far from having exhausted the potentialities of model building UK and further insights can be expected from modelling TPA (which is presently under way), crucial features of UK are left to be detected by experiment. An X-ray crystallographic study would be even more exciting if applied to its high molecular mass form.

REFERENCES

[1] Günzler, W.A., Steffens, G.J., Ötting, F., Buse, G. and Flohé, L. (1982) Hoppe-Seyler's Z. Physiol. Chem. 363, 133–141.
[2] Steffens, G.J., Günzler, W.A., Ötting, F., Frankus, E. and Flohé, L. (1982) Hoppe-Seyler's Z. Physiol. Chem. 363, 1043–1058.
[3] Günzler, W.A., Steffens, G.J., Ötting, F., Kim, S.-M.A., Frankus, E. and Flohé, L. (1982) Hoppe-Seyler's Z. Physiol. Chem. 363, 1155–1165.
[4] Heyneker, H., Holmes, W., Shepard, H.M., Seeburg, P., Hayflick, J., Ward, C., Vehar, G., Steffens, G., Günzler, W.A., Ötting, F. and Flohé, L. (1983) Proc. 4th Int. Symp. Genetics of Industrial Microorganisms, Kyoto, pp.214–221.
[5] Pennica, D., Holmes, W.E., Kohr, W.S., Harkins, R.N., Vehar, G.A., Ward, C.A., Bennett, W.F., Yelverton, E., Seeburg, P.H., Heyneker, H.L., Goeddel, D.V. and Collen, D. (1983) Nature 301, 214–221.
[6] Walton, P.L. (1967) Biochim. Biophys. Acta 132, 104–114.
[7] Dayhoff, M.O. (1972) Atlas of Protein Sequence and Structure, vol.5, National Biomedical Research Foundation, Washington DC.
[8] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. 47, 45–148.
[9] Tickle, I.J. (1982) unpublished.
[10] Iones, T.A. (1978) J. Appl. Crystallogr. A 34, 863.
[11] Hermans, J. and McQueen, J.E. (1974) Acta Cryst. A 30, 730–739.
[12] Bernstein, F.C., Koetzle, T.F., Williams, G.S.B., Meyer, E.F. jr, Brice, M.D., Rodgers, S.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[13] Fehlhanner, H. and Bode, W. (1975) J. Mol. Biol. 98, 693–717.
[14] Ruehlmann, A., Kukla, D., Schwager, P., Bartels, K., Huber, R., Kohl, W., Deisenhofer, J., Steigemann, W., Kohl, W. and Ryan, C.A. (1973) J. Mol. Biol. 77, 417–436.
[15] Hartley, B.S. (1970) Phil. Trans. Roy. Soc. Ser. B 2, 257, 77–87.
[16] Greer, J. (1981) J. Mol. Biol. 153, 1027–1042.
[17] Steitz, F.A., Henderson, R. and Blow, D.M. (1969) J. Mol. Biol. 46, 337–348.
[18] Greer, J. (1981) J. Mol. Biol. 153, 1043–1053.
[19] Furie, B., Bing, D.H., Feldmann, R.J., Robison, D.J., Burnier, J.P. and Furie, B.C. (1982) J. Biol. Chem. 257, 3875–3882.
[20] Blow, D.M., Wright, C.S., Kukla, D., Ruhlmann, A., Steigemann, W. and Huber, R. (1972) J. Mol. Biol. 69, 137–144.