Journal of Computer-Aided Molecular Design, 2 (1988) 7–14
ESCOM

J-CAMD 027

Prediction of the three-dimensional structure of the enzymatic domain of t-PA

A. Heckel and K.M. Hasselbach

Department of Chemical Research, Dr. Karl Thomae GmbH, Birkendorfer Strasse 65, D-7950 Biberach 1, F.R.G.

Received 7 October 1987
Accepted 12 November 1987

Key words: Modeling; Serine protease; Computer graphics; Fibrinolysis

SUMMARY

Tissue plasminogen activator (t-PA), an enzyme of the fibrinolytic system, is responsible for lysis of fibrin via activation of plasminogen, and therefore for degradation of blood clots. There are currently no X-ray crystal structure data of the t-PA molecule available either in whole or in part. We therefore predicted the three-dimensional structure of the protease domain by means of computer-graphical methods*.

The model obtained forms a basis for understanding the binding of plasminogen to the active site of t-PA. In addition, the interactions of various inhibitors with t-PA were studied by modeling them into the active site. The model also yields an explanation for the observed amidolytic activity of t-PA in the single chain form.

INTRODUCTION

The glycoprotein tissue plasminogen activator (t-PA) plays a central role in fibrinolysis. t-PA cleaves the peptide bond Arg 560–Val 561 of plasminogen, thus converting it to plasmin. Plasmin itself then opens blocked vessels by lysing fibrin of intravascular thrombi [1,2].

Whereas earlier t-PA was available only in small amounts by isolation from melanoma cell cultures it is now produced by biotechnological methods [3]. t-PA is a protein consisting of 527 amino acids, 3 of which are potential glycosylation sites; its molecular weight is about 65 kDa.

The t-PA molecule is composed of various discrete regions [4]. The fibronectin homologous part (amino acids 1–43) plays an essential role in binding to fibrin [5]. The role of the growth factor sequence (amino acids 44–91) is not yet clear. The two kringle regions (amino acids 92–177, and 178–262), of which homologous structures were found in prothrombin [6] and plasminogen, have been proposed to be involved in the binding of t-PA to fibrin. Kringle 2 was shown to be im-

*Presented in part at the XIth International Congress on Thrombosis and Haemostasis, Brussels, 6–10 July, 1987.

portant in increasing the affinity of plasminogen to t-PA in the presence of fibrin. This mechanism results in an improved lysis of fibrin, which is much more favored than the systemic lysis of fibrinogen.

The enzymatic activity of t-PA is located in the serine protease domain (amino acids 263–527).

Up to now no X-ray structural analysis of t-PA has been reported. To gain a deeper insight into the enzymatic activity of t-PA on a molecular level, the knowledge of its three-dimensional structure is necessary. We performed modeling by means of computer-graphical methods, based on the hypothesis that homologous serine proteases exhibit similar tertiary structures. Based on the same hypothesis, models of the three-dimensional structures of thrombin, factor X [7], urokinase [8] and other proteases have been proposed [10].

Our procedure was to set up the three-dimensional structure of the t-PA serine-protease model in three steps:

(1) Alignment of the t-PA amino acid sequence with other serine proteases of known three-dimensional structure and choosing one of them as a suitable basic structure.

(2) Changing the differing side chains according to the t-PA sequence.

(3) Modeling of insertions and deletions.

METHODS

The coordinates of known structures of serine-proteases were taken from the Brookhaven Protein Data Bank. For modeling the t-PA structure, the computer programs MOLEDT (Biosym Inc.), CHEMGRAF (Chemical Design) and MOGLI (Evans & Sutherland) were used besides our own in-house software. Calculations were performed on a VAX-11/750 computer system and the structures were displayed on PS-350 and SIGMEX 6130 devices.

MODELING THE t-PA STRUCTURE

1. Alignment

The amino acid sequence of the enzymatic part of t-PA is shown in Table 1 (chymotrypsinogen numbering). The t-PA sequence was aligned with chymotrypsinogen, trypsin and elastase using the program IDEAS (NCI). On the basis of the high homology with t-PA and due to the reasonable resolution of the structure, elastase was chosen as a basis to model t-PA.

2. Replacing the side chains of elastase by those of t-PA

Using MOLEDT the differing side chains of elastase were replaced by those needed to build up the t-PA structure. The replacement was performed to maintain the side-chain orientation whenever possible. If a new side chain extended further than its precursor, standard torsion angles were used for the extended part. In this way 236 amino acids of a total of 265 were modeled.

During this work it was observed that if a particular t-PA side chain turned out to be shorter than its precursor in elastase, another interacting amino acid was found to be extended, thus maintaining hydrogen bonds or salt bridges. Therefore, this positioning of side chains seems to be of relatively high reliability.

TABLE 1

ALIGNMENT OF THE AMINO ACID SEQUENCES OF t-PA, ELASTASE, CHYMOTRYPSINOGEN AND TRYPSIN (chymotrypsinogen numbering used)

	Sequence
	1 10 20 30 36a b c d e 40
t-PA	C G L R Q Y S Q P Q F R I K G G L F A D I A S H P W Q A A I F A K H R R S P G E R F
Elastase	V V G G T E A Q R N S W P S Q I S L Q Y R S G S S W A H
Chymotrypsinogen	C G V P A I Q P V L S G L S R I V N G E E A V P G S W P W Q V S L Q D K T G F H
Trypsin	I V G G Y T C G A N T V P Y Q V S L N S G Y H
	50 61a b c d 70 80
t-PA	L C G G I L I S S C W I L S A A H C F Q E R F P P H H L T V I L G R T Y R V V P G E E E Q
Elastase	T C G G T L I R O N W V M T A A H C V D R E L T F R V V V G E H N L N Q N N G T E Q
Chymotrypsinogen	F C G G S L I N E N W V V T A A H C G V T T S D V V V A G E F D Q G S S S E K I Q
Trypsin	F C G G S L I N S Q W V V S A A H C Y K S G I Q V R L G E D N I N V V E G N E Q
	90 100 110 120
t-PA	K F E V E K Y I V H K E F D D D T Y D N D I A L L Q L K S D S S R C A Q E S S V V R T V
Elastase	Y V G V Q K I V V H P Y W N T D D V A A G Y D I A L L R L A Q S V T L N S Y V Q L G
Chymotrypsinogen	K L K I A K V F K N S K Y N S L T I N N D I T L L K L S T A A S F S Q T V S A V
Trypsin	F I S A S K S I V H P S Y N S N T L N N D I M L I K L K S A A S L N S R V A S I
	130 140 150 160
t-PA	C L P P A D L Q L P D W T E C E L S G Y G K H E A L S P F Y S E R L K E A H V R L Y P S S
Elastase	V L P R A G T I L A N N S P C Y I T G W G L T R T N G Q L A Q T L Q Q A Y L P T V D Y A
Chymotrypsinogen	C L P S A S D D F A A G T T C V T T G W G L T R Y T N A N T P D R L Q Q A S L P L L S N T
Trypsin	S L P T S C A S A G T Q C L I S G W G N T K S S G T S Y P D V L K C L K A P I L S N S
	170 180 188a b c d e f g 190 200
t-PA	R C T S Q H L L N R T V T D N M L C A G D T R S G G P Q A N L H D A C Q G D S G G P L V C L
Elastase	I C S S S S Y W G S T V K N S M V C A G G D G V R S G C Q G D S G G P L H C L
Chymotrypsinogen	N C K K Y W G T K I K D A M I C A G A S G V S S C M G D S G G P L V C K
Trypsin	S C K S A Y P G Q I T S N M F C A G Y L Q G G K D S C Q G D S G G P V V C S
	210 220 230 240
t-PA	N D G R M T L V G I I S W G L G C G Q K D V P G V Y T K V T N Y L D W I R D N M R P
Elastase	V N G Q Y A V H G V T S F V S R L G C N V T R K P T V F T R V S A Y I S W I N N V I A S N
Chymotrypsinogen	K N G A W T L V G I V S W G S S T C S T S T P G V Y A R V T A L V N W V Q Q T L A A N
Trypsin	G K L Q G I V S W G S G C A Q K N K P G V Y T K V C N Y V S W I K Q T I A S N

3. Modeling of insertions and deletions

On the basis of the alignment some amino acid sequences of t-PA were not contained within the sequence of elastase, but could be obtained from the known structures of other serine proteases such as trypsin, chymotrypsin or chymotrypsinogen. To model these insertions, five corresponding fragments of these enzymes were cut out and inserted into the t-PA model at those positions where information was missing from the elastase structure. For example, the sequence of the amino acids 1–12 (Table 1) was taken from chymotrypsinogen (by accordingly changing differing side chains). The position of this fragment on the surface of t-PA was adjusted, relative to the elastase part by fitting its carboxy terminal part to the first residue of elastase and by allowing a disulfide bridge between residues Cys 1–Cys 122 (chymotrypsinogen numbering is used throughout).

Besides the insertions mentioned, there are additional amino acids contained in the sequence of elastase which are not part of the t-PA sequence (based on our alignment). For example, the amino acids Ala 99a–Ala 99b are not contained in t-PA. To model this deletion the sequence of residues Asn 95–Ile 103 of trypsin was cut out and aligned with the corresponding positions of residues Asn 95–Ile 103 of elastase, which were thus replaced.

Further examples were the positions of Ala 146 and Asp 184 which were modeled as insertions according to the trypsin structure and the deletions Leu 217a–Gly 218. Interestingly, all insertions and deletions are located on the surface of the t-PA structure. The model obtained so far included about 95% of the amino acids at their correct position.

At this stage, there were still 15 amino acids of t-PA for which there was no homologous structure available from other serine proteases. All these additional residues are located on the surface of the current t-PA model and were introduced into the t-PA structure as β-turns according to our secondary structure prediction based on Chou-Fasman rules. For example, the amino acids His 36a–Pro 36e are not found in any other serine-protease. Therefore this sequence was modeled as a β_{II}-turn, and inserted by adjusting its end positions to the corresponding residues in the t-PA model (by neglecting the corresponding elastase residues). Analogously, the sequences Arg 61a–

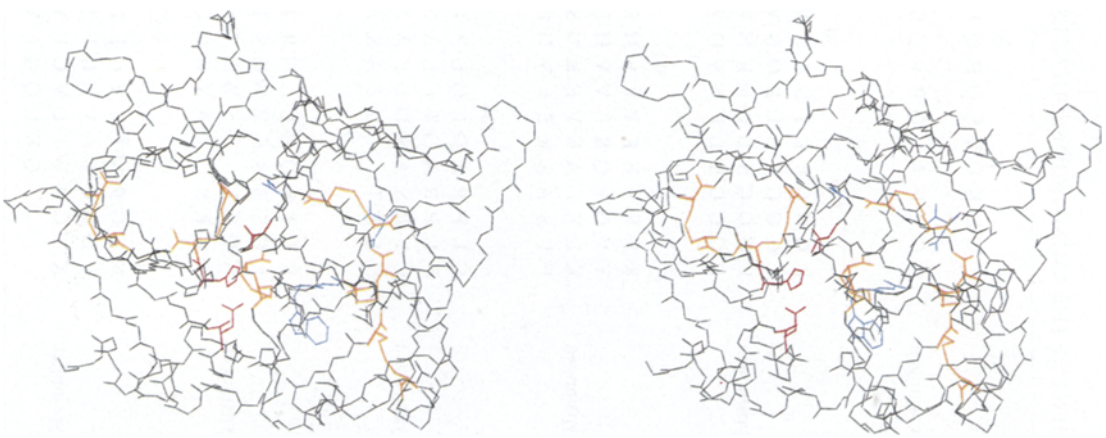

Fig. 1. Stereophotograph of the serine-protease part of t-PA (black: backbone; yellow: disulfide bridges; red: essential amino acids of the catalytic triad; and blue: other amino acids of the active site).

Pro 61d, Ser 109a–Ser 109d and Pro 188b–His 188g were introduced, yielding a model containing the full sequence of the serine protease part of t-PA. To optimize this model all torsion angles of the side chains were adjusted in order to avoid van der Waals' contacts and to form hydrogen bonds and salt bridges.

In order to predict the interaction of Arg 560 of plasminogen or t-PA inhibitors with the active site of t-PA, published structural information on protease–inhibitor complexes was used. Whereas a number of such complexes containing lysine in the active site have been investigated, the structures of which are available from the Brookhaven Protein Data Bank, no protease–inhibitor complexes with arginine in the active site have been resolved by X-ray analysis. However, a structure of a trypsin–benzamidine complex is documented in which the positively charged amidinium group forms a salt bridge with Asp 189 of trypsin. From this structural information, we set up a model of an arginine–t-PA complex in which (as in the case of benzamidine) the guanidinium group of arginine is postulated to be strongly bound by coulombic interactions to Asp 189.

RESULTS AND DISCUSSION

The structure of our t-PA-model is shown in Fig. 1. Similar to elastase, the model contains four cysteine bridges Cys 42–Cys 58, Cys 136–Cys 201, Cys 168–Cys 182 and Cys 191–Cys 220 and, in addition, the two disulfide bridges between Cys 1–Cys 122 and Cys 50–Cys 111. The model therefore predicts six cysteine bridges in the serine protease part of t-PA. An indication of the reliability of the model is the fact that nearly all charged side chains of residues such as Lys, Arg, Asp and

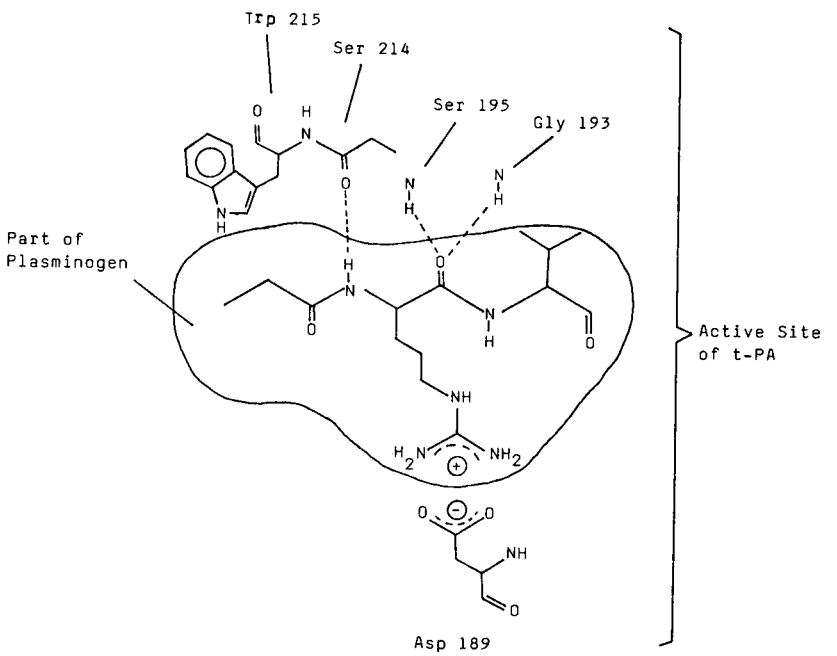

Fig. 2. Binding interactions of Arg 560 of plasminogen with the active site of t-PA.

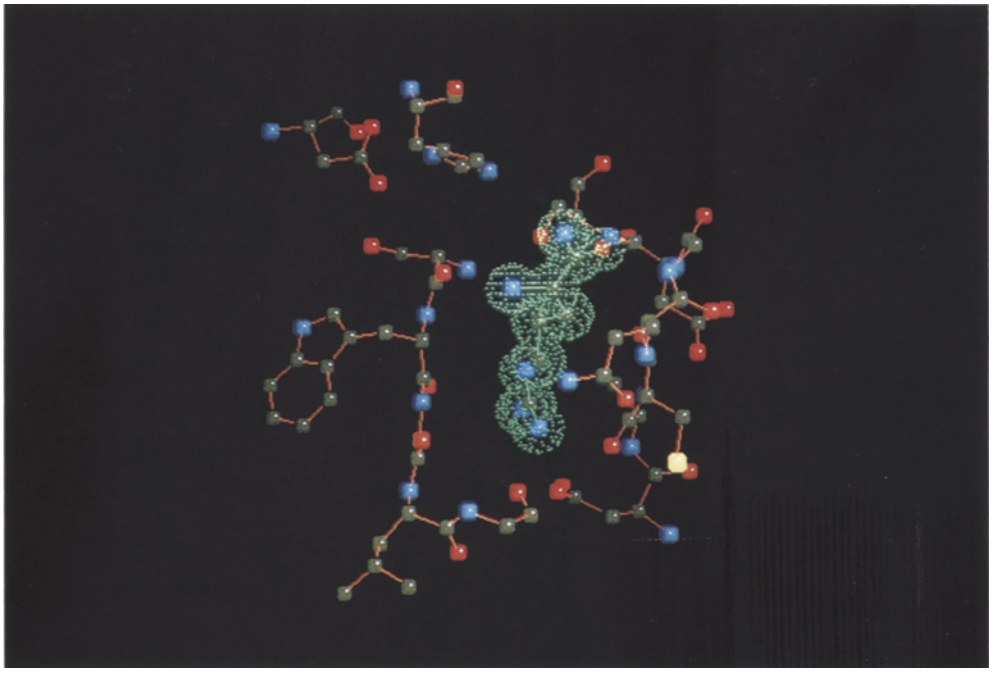

Fig. 3. Active site of t-PA binding Arg 560 of plasminogen.

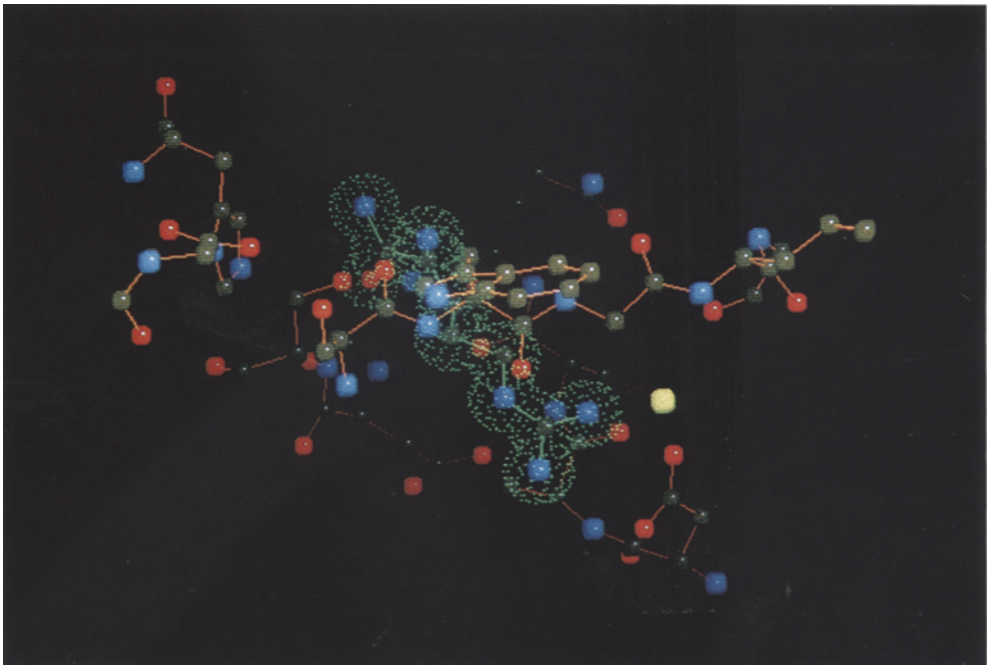

Fig. 4. Active site of t-PA binding Arg 560 of plasminogen, a different view from that shown in Fig. 3.

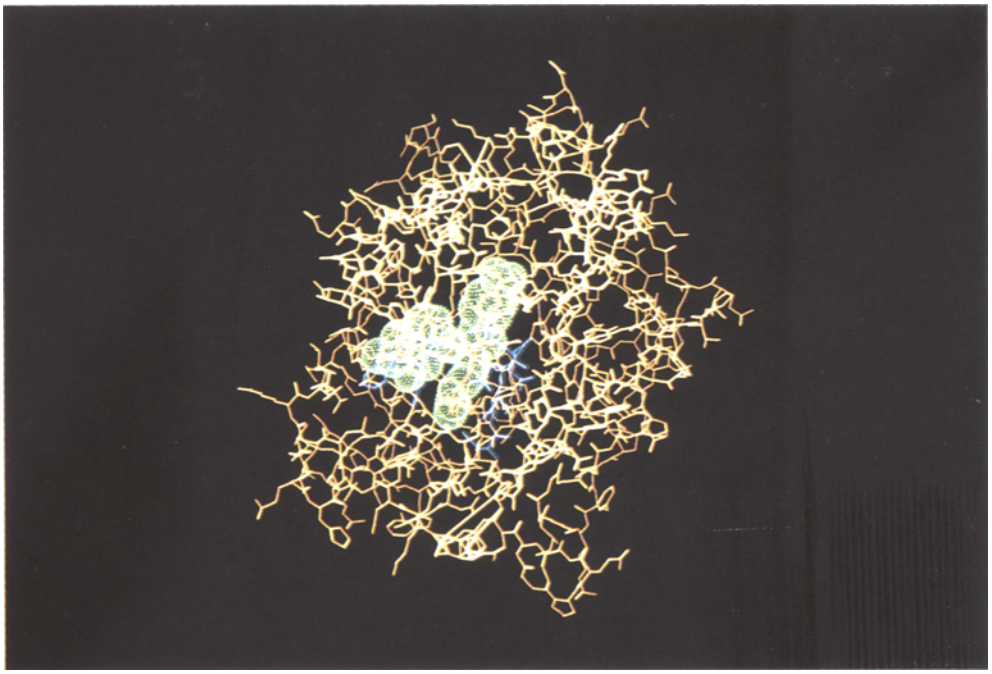

Fig. 5. Model of the protease domain of t-PA (yellow) and the active site of t-PA (blue) binding the chromogenic substrate S 2228 (orange structure, surface: dotted, green).

Glu are positioned near the surface, pointing into the aqueous phase, whereas the other hydrophobic residues form the core. Accordingly Asn 175, which is glycosylated in the native enzyme, is located at the surface.

The residue Lys 143 is located near the active site and could be assumed to form a salt bridge to Asp 194. Lys 143 thereby could stabilize the conformation of the active site. This may be the reason why t-PA, in contrast to other serine proteases, does not need to be activated by cleaving a bond between residues 15–16 [9], but posseses amidolytic activity in both the one-chain and two-chain forms [10].

Besides those features already mentioned, the model of the serine protease part of t-PA is of particular interest with respect to questions concerning the binding of its substrate plasminogen and t-PA inhibitors of lower molecular weight. Plasminogen is activated by t-PA via cleavage of the peptide bond Arg 560–Val 561, whereby Arg 560 is bound in the active site of t-PA.

According to our model, the guanidino group of Arg 560 interacts with Asp 189. In addition the carbonyl group of arginine is fixed in space by two hydrogen bonds to the N-H groups of Gly 193 and Ser 195, respectively. A third hydrogen bond is formed between the N-H group of arginine and the carbonyl group of Ser 214. These four interactions (Fig. 2) fix arginine to form an extended structure in the active site of t-PA (Figs. 3 and 4).

The chromogenic substrate S 2288 (D-Ile-Pro-Arg-NH-*p*-nitroanilide) and the irreversibly binding inhibitor PPACK (D-Phe-Pro-Arg-CH_2-Cl) were also modeled into the active site of t-PA (Fig. 5). Besides those interactions mentioned in the case of arginine, there is additional binding

between the N-H group of D-Phe (and D-Ile, respectively) to the carbonyl group of Gly 216. This hydrogen bond forces the side chain of D-Phe (D-Ile) into the direction of the aromatic ring of Trp 215 giving rise to an additional hydrophobic interaction between the aromatic ring systems, which could not be obtained on replacing the D-amino acids D-Phe (D-Ile) by their natural L-counterparts. Our t-PA model therefore explains the high affinity of S 2288 and PPACK and the role of the D-amino acids contained in these inhibitors.

ACKNOWLEDGEMENT

We thank Mr. E. Kuppinger for technical assistance.

REFERENCES

1 Pery, R.S., Drugs of the Future, 10 (1985) 835–836.
2 Collen, D., Drugs, 31 (1986) 1–5.
3 Pennica, D., Holmes, W., Kohr, W.J., Harkins, R.N., Vehar, G.A., Ward, C.A., Bennett, W.F., Yelverton, E., Seeburg, P.H., Heyneker, H.L. and Goeddel, D.V., Nature, 301 (1983) 214–227.
4 Ny, T, Elgh, F. and Lund, B., Proc. Natl. Acad. Sci. U.S.A., 81 (1984) 5355–5359.
5 van Zonneveld, A.J., Veerman, H. and Pannekoek, H., J. Biol. Chem., 261 (1986) 14214–14218.
6 Park, C.H. and Tulinsky, A., Biochemistry, 25 (1986) 3977–3982.
7 Furie, B., Bing, D.H., Feldmann, R.J., Robinson, D.J., Burnier, J.P. and Furie, B.C., J. Biol. Chem., 257 (1982) 3875–3882.
8 Strassburger, W., Wollmer, A., Pitts, J.E., Glover, I.D., Tickle, I.J., Blundell, T.L., Steffens, G.J., Günzler, W.A., Ötting, F. and Flohe, L., FEBS Lett., 157 (1983) 219–223.
9 Huber, R. and Bode, W., Acc. Chem. Res., 11 (1978) 114–122.
10 Tate, K.M., Higgins, D.L., Holmes, W.E., Winkler, M.E., Heyneker, H.L. and Vehar, G.A., Biochemistry, 26 (1987) 338–343.
11 Blundell, T. and Sternberg, M.J.E., Trends Biotechnol., 3 (1985) 228–235.