The three-dimensional structure of thymidine kinase from Herpes simplex virus type 1

Klemens Wild[a], Thomas Bohner[b], André Aubry[c], Gerd Folkers[b], Georg E. Schulz[a,*]

[a]*Institut für Organische Chemie und Biochemie, Albert-Ludwigs-Universität, Albertstr. 21, Freiburg im Breisgau, Germany*
[b]*Department Pharmazie, Eidgenössische Technische Hochschule, Zürich, Switzerland*
[c]*LCM3B, Université Henri Poincaré, Nancy I, Vandoeuvre Cedex, France*

Received 26 May 1995; revised version received 8 June 1995

Abstract Recombinant thymidine kinase from Herpes simplex virus type 1 (ATP:thymidine 5′-phosphotransferase; EC 2.7.1.21), an enzyme of therapeutic importance, was purified and crystallized in an N-terminally truncated but still fully active form. The three-dimensional structure was solved by X-ray diffraction analysis at 3.0 Å resolution using isomorphous replacement. The chain fold is presented together with the bound substrates thymidine and ATP. Three chain segments at the surface could not be located. The chain fold, the location of the substrates and presumbly also the catalytic mechanism resemble the well-known adenylate kinases.

Key words: Thymidine kinase; Herpes simplex virus; X-Ray analysis; Phosphoryl transfer; Chain fold

1. Introduction

Thymidine kinase (TK) is the key enzyme in the pyrimidine salvage pathway. Whereas the mammalian cytosolic isozyme catalyzes the transfer of the γ-phosphate from ATP·Mg^{2+} to thymidine to produce thymidine monophosphate (dTMP), the virus encoded isozyme repeats this step yielding thymidine diphosphate (dTDP) [1,2]. Moreover, host TK accepts only pyrimidines, whereas viral TK also accepts purines, and even those carrying an acyclic ribose analogue such as the prodrugs acyclovir or gancyclovir. In virus-infected cells these prodrugs are activated to their triphosphates and incorporated into the viral DNA resulting in chain termination [3]. As the activation is catalyzed nearly exclusively by the viral TK, only infected cells are hit [4]. Moreover, TK is used in the so-called 'virally directed enzyme prodrug therapy', where the gene for viral TK is introduced to tumor cells followed by administration of gancyclovir. The viral TK renders the proliferating cells susceptible to this treatment by transforming gancyclovir into a cytotoxic drug.

The structure of the viral TK is expected to provide a basis for rational design of more efficient substrates, inhibitors or enzymes for severe Herpes viral infections and for the gene therapy approach on tumors. Here, we describe the crystal structure of a recombinant viral TK from Herpes simplex virus type 1 (HSV1) expressed as fusion protein in *E. coli*. The enzyme is N-terminally truncated by 33 residues, but still fully active. It is dimeric with $M_r = 2 \times 37{,}129$.

2. Materials and methods

2.1. Materials

Enzymes used in molecular cloning were obtained from Boehringer-Mannheim. Thrombin, glutathione-agarose, ATP-agarose and other chemicals were purchased from Sigma. The plasmid pGEX2T was purchased from Pharmacia. The plasmid pBR322-TK containing the gene for TK from HSV1 was a gift from S. McKnight (Baltimore). *E. coli* strain 71/18 was the host for cloning procedures. The TK-deficient *E. coli* strain KY895 was used for expression of the fusion protein glutathione-*S*-transferase-TK.

2.2. Expression and purification

The bacterial expression vector pGEX2T-TK was constructed as described earlier [5]. Purification of the glutathione-*S*-transferase-TK fusion protein was carried out by glutathione affinity chromatography, thrombin cleavage followed by ATP affinity chromatography as described [2,6]. Expression and purification were monitored by SDS-PAGE using the Phast System from Pharmacia; the enzyme activity was measured as described [6]. Protein concentrations were determined according to Bradford [7] and by amino acid analysis.

2.3. Crystallization

For crystallization, we used the hanging drop method at 4°C. The purified N-terminally truncated protein started at residue position 34, where the viral TK is sensitive to thrombin cleavage. The truncation was confirmed by N-terminal sequence analysis. The enzyme was kept in a buffer containing 10 mM MOPS at pH 7.5, 1 mM EDTA, 0.1% Triton X-100, 2 mM DTT, 50 μM thymidine and 10 mM ATP. Using a centricon the protein was concentrated to 40 mg/ml and subsequently diluted with the same buffer (but without Triton X-100 and ATP) to 5 mg/ml. The resulting protein solution was mixed in a 1:1 ratio with the crystallization buffer (40–50 mM KH_2PO_4, pH 5.5, 1 mM DTT, 50 μM thymidine, 15–20% PEG-8000, 0.02% NaN_3) to give a drop of 5 μl. The reservoir contained the crystallization buffer.

Crystals appeared after 3 days and grew within 1–2 weeks to sizes of about $300 \times 200 \times 200$ μm^3. They belong to space group $I4_1$ with $a = b = 83.4$ Å, $c = 156.7$ Å and one subunit of the dimeric enzyme in the asymmetric unit. At 67% the solvent content is rather high. The crystals are reasonably stable in the X-ray beam and diffract to 2.8 Å resolution under X-radiation from a rotating anode and to 2.5 Å at a synchrotron (EMBL-outstation, DESY, Hamburg). They differ from crystals of TK from HSV1 reported earlier [8].

2.4. Data collection and phase determination

Data were collected using a rotating anode (model RU200B, Rigaku) together with a multi-wire area detector (model X1000, Siemens). Data processing was with program XDS [9]. Heavy atom derivatives were produced by soaking (Table 1). The difference Pattersons of all 3 derivatives were clearly interpretable. Unfortunately, all of them were single site derivatives with mercury bound to Cys-336. The heavy atom parameters were refined with program PHASIT from the PHASES package (W. Furey, 1991; University of Pittsburgh, PA, USA) yielding an essentially single isomorphous replacement electron density map. This map was poor, but could be strongly improved by solvent-flattening using program BNDRY of PHASES. The resulting map allowed us to

*Corresponding author. Fax: (49) (761) 203-5987.
E-mail: schulz@bio2.chemie.uni-freiburg.de

Abbreviations: DTT, dithiothreitol; HSV1, Herpes simplex virus type 1; PEG, polyethylene-glycol; TK, thymidine kinase.

SSDI 0014-5793(95)00680-X

place the polypeptide chain unambiguously, except for three segments at the surface. Model refinement has been started.

3. Results and discussion

3.1 Chain fold

The current model consists of 2154 non-hydrogen atoms and both substrates, thymidine and ATP. Model refinement without water has begun, the present *R*-factor is 28.9%. All of the residues lie in allowed, and 71% in most favoured regions of the Ramachandran plot (program PROCHECK in the XPLOR package). The model includes 284 of the 343 residues of the N-terminally truncated TK. Residues 34–49, 146–151 and 251–287 were not yet modeled because of low density in the initial solvent-flattened s.i.r. map. Still, the chain connectivity is unique. The lacking defined conformation of residues 34–49 and 146–151 may be caused by the truncation of residues 1–33. The secondary structures could be clearly assigned and the sequence could be well fitted to the density.

The overall fold of TK is illustrated in Fig. 1. The core of the protein contains a parallel 5-stranded β-sheet (b5, b4, b1, b3 and b2) in the same arrangement as in the adenylate kinases [10]. The sheet is surrounded by 12 α-helices (a1 to a12) covering both sides. Thus, TK belongs to the α/β proteins. Several α-helices (e.g. a1, a2, a7, a8, and a11) have equivalent counterparts in the adenylate kinases. Between helices a8 and a9 the conformation of 37 residues is not yet established due to lack of appropriate density. Using program PREDICT (package HUSAR, DKFZ, Heidelberg) with 4 current prediction methods no secondary structure is suggested for this chain segment. Here, the chain appears to form an extra domain consisting of the omitted residues together with helices a9, a10 and loop a9–a10. As there is not much density, this domain may well be mobile.

The subunit arrangement in the dimer is depicted in Fig. 2.

Table 1
Data collection and phase determination by isomorphous replacement

	Native	Derivatives			
		HMSA[a]	PCMPS-1[b]	PCMPS-2[b]	MC[c]
Resolution [Å]	3.0	3.5	3.0	3.5	3.2
Number of crystals	1	1	1	1	1
Completeness [%]	80	63	76	80	81
Outermost shell [%]	58	51	76	49	61
R_{sym} [%][d]	9.1[e]	8.2	11.9	8.7	8.8
R_{nat} [%][f]		17	31	18	16
Heavy atom sites					
x		0.085	0.085	0.084	0.085
y		0.109	0.112	0.110	0.105
z		0.750	0.750	0.750	0.750
Rel. occupancy		77	61	62	81
Phasing power		1.6	1.4	1.6	1.4

[a] The soak was with 1 mM 3,5-(hydroxymercuri)salicylic acid for 1 day.
[b] Both soaks were with 1 mM *p*-(chloromercuri)-phenyl-sulfonic acid for 3 days. Labels 1 and 2 refer to data collection with synchrotron radiation (EMBL-outstation, Hamburg) and rotation anode, respectively.
[c] The soak was with 1 mM mercurichrome for 2 days.
[d] $R_{sym} = \Sigma_{h,i}|I(h)_i - < I(h) > |/\Sigma_{h,i} I(h)_i$.
[e] The value corresponds to all collected data to 2.9 Å resolution as given by program MERGE (B. Dijkstra, Groningen, The Netherlands). The outermost shell 3.0–2.9 Å was of low quality and has been discarded.
[f] $R_{nat} = 2 \cdot \Sigma|F_{deriv} - F_{nativ}|/\Sigma(F_{deriv} + F_{nativ})$.

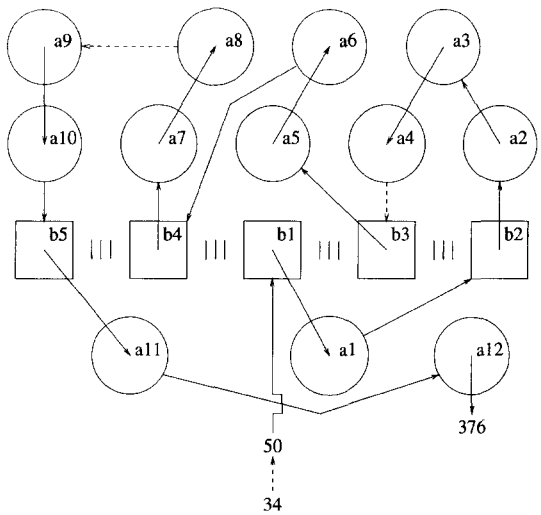

Fig. 1. Sketch of the chain fold of thymidine kinase from Herpes simplex virus type 1. To avoid crossovers the sketch is somewhat simplified rendering some helix positions disputable. All β-strands (quadrangles) run towards the viewer, the orientations of the helices (circles) are indicated by the arrows. Dashed lines denote chain segments that are not yet modeled. The central residues of secondary structure elements b1, b2, b3, b4, b5, a1, a2, a3, a4, a5, a6, a7, a8, a9, a10, a11, and a12 are 52, 79, 159, 203, 326, 67, 87, 103, 131, 172, 190, 215, 239, 292, 314, 340, and 364, respectively.

Helices a4 and a6 are at the center of the interface. Furthermore, the interface includes loop a2-a3, part of loop a9-a10, and the ends of helices a10 and a12. As illustrated by appropriate model orientation in Fig. 2, the interface is large and almost planar. It contains a twofold crystallographic axis. Thymidine binds closer to the dimer interface than ATP.

3.2 Active center

The active center is defined by the bound substrates; a close-up view is given in Fig. 3. In spite of the presence of both substrates, no reaction occurs because EDTA in the buffer removes the required Mg^{2+}. A crucial part of the active center is the glycine-rich loop connecting β-strand b1 with helix a1. This loop contains the sequence fingerprint -G-X-X-G-X-G-K-T- forming a giant anion hole [11] that accomodates the β-phosphoryl group of ATP (Fig. 3) like in the adenylate kinases [12] and in numerous other nucleotide binding proteins [13]. As deduced from the adenylate kinases the lysine of this fingerprint is essential for phosphoryl transfer during catalysis [12].

The substrate thymidine has clear density in a pocket formed by helices a3, a4 (kinked at Pro-131) and a5. Its ribose moiety points towards the glycine-rich loop. The tight binding corresponds to the low K_m-value of 0.2 μM for this nucleoside [2]. In front of the ribose moiety of thymidine (view of Fig. 3) there is space for an additional phosphoryl group, such that also dTMP can be accomodated as substrate. ATP binds at the giant anion hole and extends along helix a7 (Fig. 3). Its binding site is much more open than the thymidine site in agreement with the much higher K_m-value of 118 μM for ATP [14]. Obviously, Mg^{2+} is not required for ATP binding in correspondence with the findings for adenylate kinases [15].

The bottom of the active site (view of Fig. 3) includes sequence motif -F^{161}-D-R-H-, which is highly conserved among

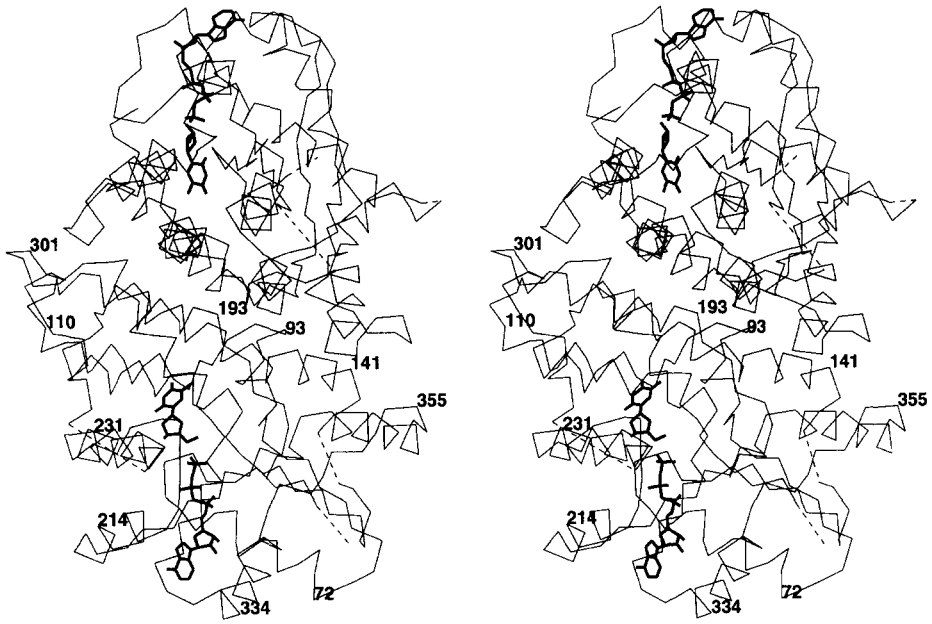

Fig. 2. Stereo-view of the current C_α-backbone of dimeric thymidine kinase from Herpes simplex virus type 1 with numbering and the bound substrates thymidine and ATP. The dashed lines indicate regions of low density which have not yet been modeled. The chosen orientation emphasizes the planarity of the interface. The two-fold axis runs from the lefthand side in the front to the right hand side in the rear.

the thymidine kinases [16,17]. It consists of the C-terminal end of β-strand b3 and part of the following loop. In the adenylate kinases the equivalent of Asp-162 fixes the Mg^{2+} required for catalysis via water molecules [18]. Furthermore, the change Asp-162→Asn in TK from HSV1 yields a completely inactive enzyme [19]. Arg-163 points towards the ribose of thymidine and is likely to bind the additional phosphoryl group of dTMP. A similarly located arginine in the adenylate kinases binds to the α-phosphate of AMP [12]. A natural mutation (to Gln) of the respective arginine in TK of Varicella zoster virus resulted in acyclovir resistance [20], corroborating its importance for the phosphorylation of dTMP.

In conclusion, we find that the presented preliminary structure in combination with the extensive knowledge about the distantly related adenylate kinases provides already a solid framework for understanding the catalysis of the viral TK.

Acknowledgements: We thank Dr. S. McKnight for providing us with the gene of TK from HSV1, Dr. J. Fetzer for helpful discussions about cloning and purification, and Dr. K.S. Wilson and his team for help in collecting one of the data sets with synchrotron radiation. The project was supported by the Land Baden-Württemberg.

References

[1] Chen, M.S., Walker, J. and Prusoff, W.H. (1979) J. Biol. Chem. 254, 10747–10753.
[2] Michael, M., Fetzer, J. and Folkers, G. (1994) Eur. J. Biochem. 226, 219–226.
[3] Coen, M.D. (1992) Semin. Virol. 3, 3–12.
[4] Elion, G.B., Furman, P.A., Fyfe, J.A., DeMiranda, P., Beauchamp, L. and Schaeffer, H.J. (1977) Proc. Natl. Acad. Sci. USA 74, 5716–5720.
[5] Fetzer, J. and Folkers, G. (1992) Pharm. Pharmacol. Lett. 2, 112–114.

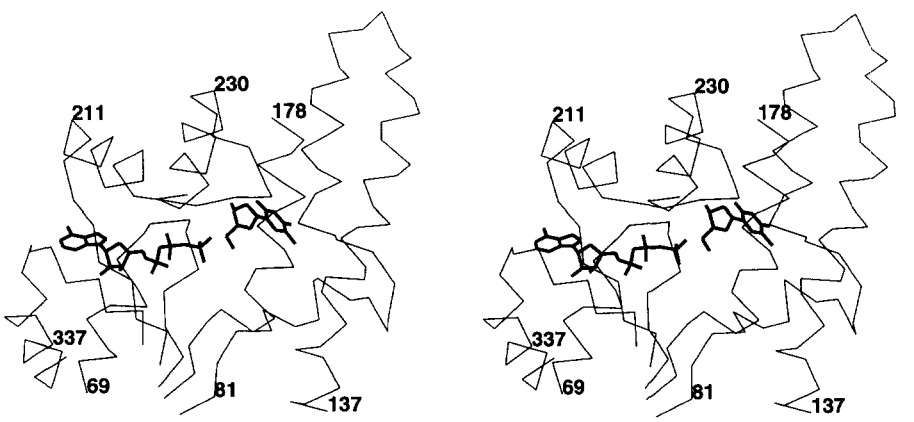

Fig. 3. Stereo-view of the C_α-backbone around the active center together with the bound substrates thymidine and ATP. Presented are five polypeptide segments, all starting at the carboxy-terminal end of the parallel 5-stranded β-sheet. The segments starting at positions 52, 80, 160, 202, and 324 contain α-helices a1, a2 & a3 and part of a4, a5, a7 and part of a8, and a10, respectively. The glycine-rich loop encircles the β-phosphoryl group of ATP.

[6] Fetzer, J., Michael, M., Bohner, T., Hofbauer, R. and Folkers, G. (1994) Protein Exp. Purif. 5, 432–441.
[7] Bradford, M.M. (1976) Anal. Biochem. 72, 248–254.
[8] Sanderson, M.R., Freemont, P.S., Murthy, H.M.K., Krane, J.F., Summers, W.C. and Steitz, T.A. (1988) J. Mol. Biol. 202, 917–919.
[9] Kabsch, W. (1988) J. Appl. Crystallogr. 21, 916–924.
[10] Schulz, G.E., Elzinga, M., Marx, F. and Schirmer, R.H. (1974) Nature 250, 120–123.
[11] Dreusicke, D. and Schulz, G.E. (1986) FEBS Lett. 208, 301–304.
[12] Müller, C.W. and Schulz, G.E. (1992) J. Mol. Biol. 224, 159–177.
[13] Schulz, G.E. (1992) Curr. Opin. Struct. Biol. 2, 61–67.
[14] Waldmann, A.S., Häusslein, E. and Milman, G. (1983) J. Biol. Chem. 258, 11571–11575.
[15] Berry, M.B., Meador, B., Bilderback, T., Liang, P., Glaser, M. and Phillips, G.N.Jr. (1994) Proteins: Struct. Funct. Genet. 19, 183–198.
[16] Folkers, G., Trumpp, S., Eger, K., Roth, H.J. and Schwoebel, W. (1987) Med. Sci. Res. 124, 137–140.
[17] Balasubramaniam, N.K., Veerisetty, V. and Gentry, G.A. (1990) J. Gen. Virol. 71, 2979–2987.
[18] Abele, U. and Schulz, G.E. (1995) Protein Sci. 4 (in press).
[19] Fetzer, J., Folkers, G., Müller, I. and Keil, G.M. (1993) Virus Genes 7, 205–209.
[20] Sawyer, M.H., Inchauspe, G., Biron, K.K., Waters, D.J., Straus, S.E. and Ostove, J.M. (1988) J. Gen. Virol. 69, 2585–2593.