Herpesviral thymidine kinases: laxity and resistance by design

James S. Evans,1 Kevin P. Lock,1 Barry A. Levine,1 J. N. Champness,2 M. R. Sanderson,2 W. C. Summers,3 P. J. McLeish4 and Alexander Buchan5

1, 5School of Biochemistry1, and Department of Infection, Medical School5, University of Birmingham, Birmingham B15 2TT, UK
2 Division of Biomedical Sciences, The Randall Institute, King’s College, Drury Lane, London WC2B 5RL, UK
3 Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA
4 Vaccine Research Foundation, c/o Department of Infection, Medical School, University of Birmingham, Birmingham B15 2TT, UK

Introduction

Many, but not all, herpesviruses induce an elevated level of thymidine kinase (TK) activity in infected cells (Morrison, 1991). The TK transcribed from the viral DNA differs in several respects from the cellular TK of the host. The TK found in uninfected cells of host species are quite substrate-specific, using Mg2+—ATP to monophosphorylate only one substrate, thymidine. The herpes simplex virus type 1 (HSV-1) TK shows a much broader range of phosphorylation activities, being a deoxyxypirimidine kinase able to phosphorylate a range of nucleoside analogues, to carry out multiple phosphorylation on thymidine and to use nucleotide triphosphates other than ATP as the phosphate source. The mutability of the viral TK enzyme also appears to be greater than that of the cellular TK enzyme. While the narrow specificity of cellular TK will deserve careful attention when crystallographic information becomes known, the existing level of crystallographic knowledge for HSV-1 TK prompted us to examine structural features, including long-range factors, that may be crucial to its function.

Most herpesviral TK molecules have a sequence length of 310–380 amino acids, although a few are shorter and some others are N- or C-terminally elongated. HSV-1 TK is 374 residues long and is thought to function as a homodimer (Waldman et al., 1983). Two reports on its structure have confirmed the expectation that a loop formed by residues 56–63, containing glycine residues, is important for ATP binding. One structural determination (Wild et al., 1995, 1997) shows thymidine/TMP and ADP bound adjacent to one another in the absence of Mg2+, while the other (Brown et al., 1995) reveals an almost identical structure containing thymidine bound with a sulphate ion in place of a phosphate source.

Knowledge about the crystal structure for a protein with distinctive characteristics tends to stimulate or intensify interest in correlating biochemical properties with structural features of the molecule, beyond the level previously possible using knowledge of primary sequence and predictive models for secondary structure. A recent interpretation of the binding of a range of nucleoside analogues to the TK of HSV-1 by De Winter & Herdewijn (1996) exemplifies such interest following closely upon the appearance of reports of partial structure determinations for HSV-1 TK complexed with natural and synthetic substrates (Brown et al., 1995; Wild et al., 1995).

Our analysis of the structure of the HSV-1 TK indicates that several non-sequential residue contacts and long-range effects appear to influence the well-known breadth of the range of nucleoside substrates, both those occurring naturally as well as those of therapeutic interest. Sequestration of the nucleoside substrate by the viral TK involves flexible van der Waals and hydrophobic contacts in a cavity bounded by a set of interacting helical segments and complemented by confining contacts from a loop segment. These contact regions are linked by a matrix of specific interactions with other areas of the molecule to enable concerted correlated adjustments in relation to the substrate. Seemingly distant point mutations that alter the relative orientations of the helices which surround the nucleoside can thereby result in substantial variances in substrate specificity, overall molecular integrity, and enzymatic parameters.

Conformational features of HSV-1 TK

We undertook an examination of the reported crystal structure of TK from HSV-1 (Brown et al., 1995; Wild et al., 1995, 1997), with a view to correlating structural information at the intramolecular level with data that had previously become known through biochemical and mutational studies. We looked for varied illustrations of crucial structural elements, e.g. any residue(s) that might have special importance for coherence of a local structural domain, or some particular group of residues from sequentially remote regions that draw one another to a locus of likely functional importance, or multiple contacts that might work together in supporting a quaternary structure. In each instance, we then considered data from sequence comparisons in order to distinguish what may only be specific to this one determined structure from what
might be usefully conjectured about viral TK molecules more generally.

Profile of the nucleoside binding site

The site that holds thymidine derives from a limited set of contacts supported by subsidiary stabilizing factors involving specific residues from several separate locations along the primary sequence. A distinctive feature of the HSV-1 TK is that the nucleoside or drug analogue substrate binding site is composed of amino acid side-chains derived primarily from a set of roughly parallel helices (Trp-88 from residues 85–89, which form the second helix in the structure; Ile-100 and Tyr-101 from 96–108, helix 3; Gln-125 and Met-128 from 114–130, specifically Tyr-172 from 171–178, helix 6; His-58, prior to helix 2; Arg-163, Ala-167 and Ala-168, prior to helix 6; Arg-222, Glu-225), as depicted in Fig. 1. These helices occur in the N-terminal half of the molecule following the glycine-rich loop involved in binding the phosphate moiety of ATP (residues 56–63, some strongly conserved; see legend to Fig. 2).

The two most nearly parallel of these helical regions, 96–108 (helix 3) and 114–130, the latter occurring within a longer helical span (helix 4) that extends to residue 142 but with a kink at Pro-131 (Wild et al., 1995), make numerous lateral van der Waals contacts (i.e. less than 4 Å) with each other (Fig. 2). A third helical region, 171–178 (helix 6), makes contacts from Arg-176 to both Ile-100 and Gln-104 in helix 3, the latter contact being especially close (Table 1). Arg-176 also makes a lateral contact via the 114–130 helix to Met-121, while Ala-168 makes another especially close contact to Gln-125 (Table 1). These several sets of lateral contacts among the three interacting helices establish the spatial parameters of the local environment for the nucleoside substrate, while still leaving some measure of laxity within the nucleoside binding region. Our analysis of the structure was therefore substantially guided by the thought that other side-chains along these helices, as well as other packing factors determining their relative arrangement in space, must also be essential for conferring distinctive binding and catalytic properties to the HSV-1 TK molecule and homologues.

Role of Cys-336

Three HSV-1 TK mutants with resistance to the drug (E)-5-(2-bromovinyl)-2′-deoxyuridine (BVdU) have been reported (Larder et al., 1983) and sequenced (Darby et al., 1986). One of these mutants, called S1, differs from wild-type TK in a single substitution of tyrosine in place of cysteine at position 336. As described below, this is perhaps the best example of the contribution of long-range conformational effects to the stability and phosphorylation efficiency of HSV-1 TK. In an external view of a space-filling model of the enzyme, Cys-336 is nearly invisible and occurs at the N-terminal end of a helical region (Gly-335 to Thr-344), which is exposed. The CB and SG atoms of Cys-336 are bordered by Tyr-329 and by the backbone loop formed by preceding residues (331–333) with the side-chain of Gln-331 hydrogen bonding to the adenine ring of the bound ADP (Wild et al., 1997) (references to specific atoms within amino acids are given using the notation of PDB files for the atoms comprising the peptide backbone linkage (N, CA, C, O) and for side-chains (Xn), where X is the atom type (C, O, N, S), Y denotes the location of the atom outward from the backbone (B, beta; G, gamma; D, delta; E, epsilon; Z, zeta; H, eta) and n denotes one of several chemically similar groups, e.g. OD1 or OD2 for the two carboxylate oxygen atoms of an aspartate side-chain). In addition, the CB and SG atoms of Cys-336 make van der Waals contacts with the methyl group of Thr-65, which follows the glycine-rich nucleotide phosphate binding loop (residues 56–63; Fig. 2). In view of the spatial relationships of the CB and SG atoms of Cys-336 to preceding residues, the introduction of the bulkier phenolic moiety of tyrosine in place of a sulphydryl group here would induce significant disruption in this part of the molecule. Local relative displacements of the magnitude of one full atomic diameter would be required in order to make room for a rigid ring and
Fig. 2. Schematic diagram showing the network of interactions within the protein framework surrounding the nucleoside/drug contacts (italicized) to enable relayed motions of the enzyme. Certain residues and groups which are important for activity are coloured as follows: yellow, Cys-336; blue, Arg-176; red, dimer interface residues Val-119, Gin-185, Asp-306, Asp-307, Trp-310 and Gin-371; and green, nucleotide binding domain. Shown below is the sequence comparison for four herpesviral TK using residue numbering as in HSV-1 TK. Primary sequences of proteins were downloaded from the SWISS-PROT databank, where the TK are indexed with identifiers KITH XXXX. Numbers between dashes indicate the number of intervening residues not shown. Asterisks denote positions that appear to be completely conserved in all herpesviral TK sequences in the SWISS-PROT database having a total length of approximately 310–380 residues. Other similarly conserved residues include: Arg-51, Tyr-53, Asp-55, Gly-56, Gly-61, Lys-62, Pro-195, Gly-200, Asn-202, Tyr-239, Leu-242, Thr-245, Thr-287, Leu-288, Trp-310 and Leu-315. Sequence homology/conservation explorations were conducted either using standard utilities (e.g. PILEUP) of the GCG software package or using MACAW (Schuler et al., 1991) on microcomputers. Sequence pattern searches were conducted over the Internet using the server of the Bioinformatics Group at the Swiss Institute for Experimental Cancer Research (ISREC).
Table 1. Selected close contacts in HSV-1 TK with (non-hydrogen) interatomic distances less than 3.5 Å

A contact, i.e. between the entry in column A and that in the same row in column B, is listed only if the adjacency is seen using data from both monomers.

<table>
<thead>
<tr>
<th>Intramonomer contacts</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg-51 NH2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg-89 CD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg-89 NE, CZ, NH2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln-104 OE1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met-121 O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala-167 O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala-168 O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys-171 O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr-172 O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro-195 O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg-212 NH1, CZ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg-216 NH1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| B                     |     |     |
| Pro-199 O             |     |     |
| Glu-95 OE1            |     |     |
| Gln-374 OE2           |     |     |
| Arg-176 NH1           |     |     |
| Gln-125 CN            |     |     |
| Tyr-172 N, CD        |     |     |
| Gln-125 NE2           |     |     |
| Ala-175 CB            |     |     |
| Arg-176 CG            |     |     |
| Arg-318 CD            |     |     |
| Asp-330 OD1           |     |     |
| Met-60 O              |     |     |

<table>
<thead>
<tr>
<th>Intermonomer contacts*</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu-91 O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly-92 O</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val-119 CG1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn-306 N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn-306 CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val-307 N, CG2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|             |     |     |
| Gln-185 CG, NE2 |     |     |
| Val-119 CG1; Ser-123 OG |     |     |
| Gln-371 CD, OE1, OE2 |     |     |
| Gln-371 OE1      |     |     |
| Gln-371 OE1      |     |     |

<table>
<thead>
<tr>
<th>Contacts with thymidine†</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr-101 OH</td>
<td>O3'</td>
<td></td>
</tr>
<tr>
<td>Met-128 CE</td>
<td>N3</td>
<td></td>
</tr>
<tr>
<td>Glu-225 OE1 or OE2</td>
<td>O3'</td>
<td></td>
</tr>
</tbody>
</table>

* Additional residues involved with less well-defined intermonomer contacts at distances 3.5–4 Å include Tyr-87, Ile-126, Met-130, Ala-133, Val-134, Leu-188, Ala-189, Ala-192, Leu-193, Tyr-305, Phe-308, Trp-310, Leu-364, Thr-367, Phe-368 and Met-372. Six water molecules in the crystal are simultaneously within 4 Å of atoms in both monomers, although not every such water molecule has a symmetric counterpart in the data set.

† Additional residues involved with contacts to thymidine at distances less than 4 Å include His-58, Glu-83, Trp-88, Ile-100, Gln-125, Arg-163, Ala-167, Ala-168, Tyr-172 and Arg-222.

distal hydroxyl group. Relay of this re-accommodation to the nucleotide triphosphate binding site appears to occur since the mutant enzyme displays much-reduced activity towards both ATP and nucleoside or drug analogues.

It is of interest, therefore, that a cysteine residue at or very near position 336 is strongly conserved in those herpesviral TK molecules having an approximate length of 310–380 residues, with the exception of that produced by varicella-zoster virus (VZV), where a smaller alanine residue occupies position 336 instead (although four cysteine residues do occur some 20–80 positions earlier). Even this substitution of alanine is noteworthy, as it would require a variation in two codons at the genetic level; few of the single-codon variations (Trp, Tyr, Phe, Arg, Ser, Gly) would seem to be very acceptable as a substitute for Cys-336 based on steric considerations. This observation of a role for Cys-336 in local domain consolidation may also help to rationalize certain unusual aspects of the anomalous behaviour of the Cys336Tyr mutant molecule, such as its reduced thermal stability and the extreme sensitivity of its TK activity to the concentration of added salt (Larder et al., 1983), which could reflect a tendency towards partial denaturation of the tertiary structure of the S1 mutant molecule under solution conditions which the wild-type TK readily withstands.

Monomer–monomer contacts in the crystal

Attention has already been drawn by Wild et al. (1995) to the planarity of the dimer interface for HSV-1 TK. We have looked at the extent and nature of contacts across the interfacial surface. Table 1 (including footnotes) lists all the residues for which any (non-hydrogen) side-chain or backbone atom in one monomeric unit lies within 4 Å of a side-chain or backbone atom in the other monomeric unit. These residues are seen to represent 15 of the possible 20 residue types in clusters that are distributed throughout the primary sequence at positions.

Fig. 3. Representation of one monomeric unit for HSV-1 TK in ribbon representation, except that all atoms which are within 5 Å from any atom in the opposing monomeric unit (not shown) are displayed as space-filling spheres in white, light grey (4 Å contact distance), or dark grey (3.5 Å contact distance). Specific residues making the shortest of these contacts are indicated (see also Table 1). Spatial proximity explorations were conducted using the Berkeley version of RasMol on microcomputers. Structural similarity explorations were conducted using the DALI server (Holm & Sander, 1993) via the Internet.
Steric constraints imposed by the five-membered proline ring are observed than reported coordinates. These have residues, with the majority of the proline residues having a graphical distribution in the HSV-1 TK (8% of the count of all the exception of a shorter (e.g. channel catfish virus TK). These variations and the contrast, Trp-310 is very strongly conserved, the exceptions even among the human herpesvirus TK sequences. Indeed, from herpesviruses that infect different vertebrate species, or monomer contacts are not well-conserved across TK sequences (i.e. its side-chain is the least mobile).

The particular residues involved in these closest monomer–monomer contacts are not well-conserved across TK sequences from herpesviruses that infect different vertebrate species, or even among the human herpesvirus TK sequences. Indeed, Asn-306 in HSV-1 becomes His-306 in HSV-2. In sharp contrast, Trp-310 is very strongly conserved, the exceptions being those herpesviral TK sequences which have chain lengths either much longer (e.g. Epstein–Barr virus TK) or much shorter (e.g. channel catfish virus TK). These variations and the relatively flexible associations that make up the intermonomer contacts suggest that closely isenergetic conformational states of the interface play a functional role by enhancing energy redistribution during the catalytic cycle.

### Stereochemistry of the proline residues

In view of the constraints on structural adaptation associated with proline residues, we looked at their topographical distribution in the HSV-1 TK (8% of the count of all residues), with the majority of the proline residues having reported coordinates. These have trans Xaa–Pro linkages with the exception of a cis Pro–Pro linkage at Pro-154. For the trans stereochemistry of Xaa–Pro linkages, which is more commonly observed than cis both in crystal structures and in solution, the steric constraints imposed by the five-membered proline ring have the further consequence of restricting the range of the conformational torsion angle \( \varphi \) at the proline (except for Xaa = glycine) to values that place the CA hydrogen atom of Xaa very near the CD protons of proline. In turn, the side-chain of Xaa has a more limited scope of conformational mobility when Xaa is followed by proline than by any other residue type.

Consider, for example, the pairing of Glu-83 and Pro-84 (totally conserved in all herpesviral TK sequences), where the glutamate side-chain forms a hydrogen bond with the 5'-hydroxyl group of the sugar of thymidine (Brown et al., 1995; Wild et al., 1997) or the surrogate group in drug analogues (De Winter & Herdeewijn, 1996). This hydroxyl group is the nucleophile that interacts with the phosphoryl group transferred to it from ATP or other nucleotide triphosphate. We can see in the structure that the totally conserved Pro-84 and Trp-88 keep the carboxylate side-chain of Glu-83 appropriately oriented to assist catalysis (i.e. by enhancing the nucleophilicity of the 5'-hydroxyl group). A water molecule, which may be required during the course of the enzymatic reaction, is in close proximity to the carboxylate moiety of Glu-83. In addition, the backbone conformation in this region brings the carbonyl group of Glu-83 into close hydrogen bonding range of the amide group of the hydrophobic residue at position 85. It thus appeared that the sequence context of Glu-83 contributes to its side-chain orientation, and it was therefore interesting to find that a database search of known protein sequences for the residue pattern Glu-Pro-Ile/Leu/Met/Val-Xaa-Tyr-Trp- [Arg/Gln] failed to find any examples of this sequence motif except in herpesvirus TK molecules having a sequence length in the range 310–380 residues.

Two non-sequential arginine residues, Arg-163 and Arg-222, are both highly conserved and also involved in positioning the fully extended side-chain of Glu-83. Arg-222 is associated with the nucleotide \( \beta \)-phosphate in the dT. ADP complex and is part of a short loop segment (residues 219–226; see below) which shows significant deviation between the otherwise similar crystal structure of the enzyme in the presence of dT or dTMP [positional root-mean-square (rms) deviation of approximately 0.6 Å for 303 superimposed CA atoms]. Arg-163 is in a strained conformation and its side-chain shifts to fix the phosphate of dTMP. The importance of Arg-163 is underscored by reports of very low TK activity for two HSV-1 double mutants Phe161Ile/Arg163His and Arg163Pro/His164Gln from random mutational studies (Black & Loeb, 1993) and for a mutant of VZV TK which has a glutamine substituted at the position equivalent to residue 163 of HSV-1 TK and which is reported to confer resistance to the drug acyclovir (ACV) (Sawyer et al., 1988).

### Role for clustered arginine residues

Arginine residues in HSV-1 TK tend to occur along the primary sequence in clusters. Moreover, the molecule rather surprisingly contains many fewer lysine residues than histidine residues. One prominent run of positive charges, spanning a...
region pointed out by Balasubramanian et al. (1990) by analogy to an arginine-rich region in adenylate kinase, goes as follows:

Asp\(^{211}\)-Arg-His-Ile-Asp-Arg\(^{216}\)-Leu-Ala-Lys-Arg\(^{220}\)-Gln-

Arg\(^{222}\)-Pro-Gly-Glu\(^{225}\)-Arg-Leu-Asp\(^{229}\). In the structure, the

underlined positively charged residues, as well as His-58, all

come within 10 Å of the nucleoside-associated Glu-225 side-

chain, while the side-chains of three aspartate residues at

positions 211, 215, and 228 are farther away. Indeed, only Glu-

83 puts any negative charge within 10 Å of the Glu-225 side-

chain. Such a net positive environment must exert a con-

siderable long-range attractive Coulombic force on the

glutamate carboxyl group. The CA-CB-CG-CD side-chain is

observed to be fully extended to 3·8 Å (the maximum for three

aliphatic bonds). The clustered positive charges would tend to

keep the Glu-225 side-chain partially oriented even in the

absence of a nucleoside substrate, helping to preform one

important aspect of the binding site. The polypeptide con-

formation from Glu-225 to Arg-226 is also extended, with its

side-chains lying on opposite sides of the backbone,

where a short carbonyl to amide distance between them

implies the existence of a strong hydrogen bond. Glu-225,

along with Tyr-101, interacts through hydrogen bonding with

the 3' hydroxyl group of the natural substrate.

Glucosamine also derives from a segment comprising residues

219–226 which loops over the nucleoside substrate even in the

absence of bound nucleotide phosphate. In the presence of

ADP, the side-chain of the strongly conserved residue Arg-

220 has swung inwards to interact with the ψ-phosphate and,

in the dTMP·ADP complex, associates with Glu-225. Across

the preceding turn from Glu-225, residue Arg-222 also lies

close to bound thymidine because of its involvement with the

nucleotide phosphate and Glu-83. Arg-222, which supple-

ments the positive charge of Lys-62 at the triphosphate

binding site of HSV-1 TK, is also conserved except in one

bovine herpesvirus TK sequence. Moreover, aspartate replaces

glutamate at position 225 only in some of the bovine herpesvirus TK sequences. The presence of a 3' hydroxyl group surrogate does not appear to be essential for binding of drugs (e.g. ACV), although all naturally occurring substrates do contain 2'-deoxy ribose rings. Similarly, hydrogen bonding interactions with the O4' ether-like atom of the substrate do not appear to exist in the structure, since the closest protein atoms are the CD1 and CG2 groups of Ile-97. The residue corresponding to position 97 in HSV-1 TK is hydrophobic in all herpesvirus TK molecules. Perhaps the absence of a hydrogen bond donor in this vicinity ensures that the contacts to 3'- and especially 5'-hydroxyl groups will be more likely to occur.

Two other arginine residues occurring earlier in the same

positively charged segment of HSV-1 TK make separate contacts to sequentially distant residues. An amino group of the strongly conserved Arg-216 makes close contact (Table 1) with the backbone carbonyl oxygen atom of Met-60 along the ATP binding loop and, together with Gln-331, helps to

sandwich the adenine ring. Such minimal contacts with the adenine base contrast with the multiple interactions which
dock the dT/nucleoside analogue and reflect the enzyme's

latitude for nucleotide triphosphates. An amino group of Arg-

212 is similarly close to the side-chain carboxyl group of Asp-

330. These interactions involving Arg-212 and Arg-216

provide two further examples of the contacts that link the ATP

binding region to other parts of the molecule (Fig. 2).

Role of Pro-195 and Pro-199 with two arginine residues

When we looked at relative mobilities as reflected in

crystallographic B-factors, we observed two arginine residues

having especially reduced side-chain mobility, Arg-51 and

Arg-318. The numerous contacts that Arg-51 and Arg-318

make with other residues are in keeping with their relatively

low overall mobilities. We were further intrigued to find that

these residues from near the N and C termini are involved,

respectively, with two proline residues from near the middle,

Pro-199 and Pro-195. One of the N-terminal groups of Arg-51

not surprisingly forms a hydrogen bond with the hydroxyl

group of Tyr-53, but the other amino group is hydrogen

bonded to the carboxyl group of Pro-199. The long arginine

side-chain is arched, and the backbone carboxyl group of Arg-

51 forms another hydrogen bond to the amide group of Asn-

202.

In a quite different fashion, Arg-318 not only makes various

expected contacts with residues along the helix of which it is

a part (residues 306–319), but also forms hydrogen bonds

between its backbone carbonyl group and the hydroxyl group

of Thr-197 and between one of its N-terminal groups and the

backbone carbonyl group of Ile-194. Furthermore, the entire

side-chain of this arginine residue lies stretched out parallel to

the ring of Pro-195 at roughly the distance expected for

extensive van der Waals contacts. These many contacts

involving Arg-318 are supported by less close contacts

involving residues 321/322 to Thr-197.

The backbone conformation from Pro-195 to Pro-199 is

extended, partly because the hydroxyl group of Thr-197 is

folded back to form a hydrogen bond with the backbone amide

group of the same residue. Overall, the interactions involving

Arg-51 and Arg-318 provide a potential means for trans-
mition of intramolecular movement near Arg-51 and Tyr-53, which precede the glycine-rich loop motif involved with ATP

binding (residues 56–63), all the way through to the dimer

surface at the helix of which Arg-318 is a part, by way of the

spacering motif extending from Pro-195 through Pro-199.

The possible importance of this quasi-mechanical linkage is

supported by the conservation of Arg-51, Tyr-53 and Pro-195

throughout herpesviral TK sequences of similar total length

(Fig. 2). Pro-199 is less well-conserved. The most frequent

substitutions observed at position 318 are lysine or cysteine

which would also seem stereochemically consistent with most

of the interactions described above. Thus, an organized matrix
of individually weak interactions emanate from the nucleotide binding loop to knot together different regions of the molecule.

Additional non-sequential constraints

Another arginine residue, Arg-176, appears to have an effect on the characteristics of the nucleoside binding site. In the crystal structure, this residue assists in positioning Tyr-101 in two ways. One terminal nitrogen atom of the side-chain of this arginine residue approaches almost within van der Waals contact distances (4.0–4.5 Å) to all the ring atoms, while the other terminal nitrogen atom makes a hydrogen bond with OE1 of Gln-104, a residue that occurs on the same helix (residues 96–108) as Tyr-101. A substitution at position 176 in the sequence could thus produce relative backbone or side-chain displacements that would interfere with a hydrogen bond from Tyr-101 to the 3'-hydroxyl group of the substrate. Another of three reported HSV-1 mutants with resistance to drug BVDU, called Tr7, was shown to result from the single substitution Arg176Gln (Darby et al., 1986). A glutamine sidechain is two links shorter than an arginine side-chain. If this shorter side-chain were to maintain the specific hydrogen bonding contact that Arg-176 forms with Gln-104, a principal constraint on the position of the tyrosine ring would be markedly altered, perhaps allowing the latter to recede from the substrate. It is also interesting that mutant Tr7 was reported not to bind deoxycytidine, an observation indicative of an altered nucleoside binding site. The details of this analysis cannot be extended to all herperviral TK molecules, since Tyr-101 is not highly conserved. The utility of the side-chain of Arg-176 in aligning the ligand-associated residue at this position is implied, however, by the total conservation of Arg-176 itself.

At an adjacent position in the sequence, the carbonyl oxygen atom of Cys-171 comes very close to the methyl sidechain of Ala-175 (Table 1). Two engineered mutants, Cys171Ser and Cys171Gly, had similar biochemical properties to the unaltered TK molecule (Inglis & Darby, 1987). On the other hand, an Ala175Thr mutant TK that was N-terminally truncated by 45 residues ran on SDS gels with a distinctly higher apparent molecular mass than the corresponding wild-type molecule (Inglis & Darby, 1987). On the other hand, an Ala175Thr mutant TK that was N-terminally truncated by 45 residues ran on SDS gels with a distinctly higher apparent molecular mass than the corresponding wild-type molecule. This helical region (residues 171–177, helix 6), one of the four that contribute groups involved in nucleoside substrate docking, is unusual because it incorporates a proline residue midway along at position 173. Positions 171, 173, and 176 are highly conserved (Fig. 2), and Pro-173 was invariant in random mutational studies conducted in this region (Munir et al., 1992). Although this proline does not interrupt overall backbone helicity, the previous Leu-170 does introduce a kink in the backbone.

Preceding helix 6 (residues 171–177) is the highly conserved Ala-167. This residue lines the substrate binding pocket and seems to be important in avoiding steric clashes (Brown et al., 1995) since several similarly conserved polar groups are all substantially closer than 4 Å to its methyl group (including the carboxylate moiety of Asp-55, the hydroxyl group of Tyr-239, and the carbonyl group of His-164). The methyl group of neighbouring Ala-168 is well away from the 5-position of thymidine, but its backbone carbonyl oxygen atom is very close to Gln-125.

A further example of the linked substructure comes from the short quasi-helical portion of HSV-1 TK between Pro-84 and Gly-92, which positions the side-chain of Arg-89 to protrude toward the molecular surface. This positively charged group is within 5 Å of not only some ring atoms of Pro-223 but also the two negatively charged side-chains of Glu-95 and Glu-374. The shortest of these contact distances is to Glu-374, which is on the C-terminal helix that provides one dimer surface contact at Glu-371 (to Asn-306 and Val-307 of the other monomer). Several backbone hydrogen bonds impart some coherence within the quasi-helical motif containing Arg-89, and this residue may be able to convey effects from interactions involving Glu-83 rather directly toward one face of the dimer surface via Glu-374. At the same time, residues Leu-91 and Gly-92 along this sequence motif make a second contact with the opposing face of the dimer interface at residue Glu-185 of the other monomer. Additionally, the proximity of Arg-89 to Pro-223 also helps to assure that Glu-225 can interact with the nucleoside substrate appropriately. Positions 89, 95 and 223 exhibit a strong degree of conservation of residue side-chain characteristics, and nearly all herpesvirus TK molecules also contain an acidic residue within 9 residues of the C terminus.

In the latter respect, it is of interest that certain natural mutations in herpesviral TK genes result in the transcription of dysfunctional protein products which are either altered in residue sequence (Hwang et al., 1994) or severely truncated in the C-terminal region (Sawyer et al., 1988). Recent work has demonstrated that all but the very last few C-terminal residues are required for enzymatic activity of Epstein–Barr virus TK (Hsu et al., 1996). Such findings would seem to be related to the role seen for Glu-371 in HSV-1 TK in forming one set of contacts at the dimer interface. In contrast, N-terminal truncations of up to 45 residues appear not to alter enzymatic activity, but instead render the viral TK molecule more susceptible to degradation in the host cellular environment.

Functional versatility is linked to structural compliance

Herperviral TK molecules have quite different primary sequence lengths, quaternary subunit associations, and biochemical profiles toward varied substrates compared to Poxviridae TK or the cellular TK of host organisms. Since HSV-
1 TK and certain other viral TK molecules can phosphorylate deoxycytidine in addition to thymidine, attention has been drawn to the fact that the channel catfish herpesviral TK bears about as much sequence homology to human cellular deoxycytidine kinase as to HSV-1 TK (Harrison et al., 1991). Similar sequence homologies have been found between HSV-1 TK and human mitochondrial deoxyguanosine kinase, which has recently been cloned and expressed (Johansson & Karlsson, 1996). Viral TK sequences have also been compared previously to the yeast thymidylate kinase sequence with comments about their possible evolutionary relationships (Robertson & Whalley, 1988). No crystal structures of these other kinases are yet available to permit direct visualization of the structural consequences of the homologies found in their primary sequences.

Since HSV-1 TK binds thymidylate not only as the product of monophosphorylation of thymidine but as a reactant that would undergo further phosphorylation, we looked briefly at related features in the structure of thymidylate synthase (Finer-Moore et al., 1996). In the latter structure, a conserved asparagine residue not only makes two much shorter hydrogen bonds with the NH3 and O4 atoms of the pyrimidine ring than are seen for conserved Gln-125 in HSV-1 TK, but this asparagine in thymidylate synthase is also held in place by an extensive network of other hydrogen bonds that are not apparent in HSV-1 TK. Additionally, our attempted structural alignments based on the DALI algorithm (Holm & Sander, 1993) showed only weak overall similarity, i.e. roughly 3-5 Å rms deviations of α-carbon positions, when HSV-1 TK was compared to guanidylate, uridylate and adenylate kinases (uridylate and adenylate kinases themselves compare at roughly 2 Å rms deviation). HSV-1 TK thus represents a rather different structure than either thymidylate synthase or adenylate kinase, with which it might be compared on account of it showing ‘thymidylate activity’ in addition to its TK activity.

The phosphoryl transfer catalysed by HSV-1 TK is known to be stereospecific, i.e. to proceed with 100% inversion as the adenylate kinase, with which it might be compared on account of their different structure than either thymidylate synthase or adenylate kinase. The Gln-125 side-chain is observed to be rotated 180° when TK structures with bound thymidine (a pyrimidine) and bound ganciclovir (a purine nucleoside) are compared (Brown et al., 1989). Some considerable flexibility of conformation for Gln-125 may be envisioned to be required for accommodating deoxycytidine because of the difference in disposition of hydrogen bonding donors and acceptors around the heterocyclic rings of these two pyrimidines.

A further aspect of the selectivity for thymidine over deoxycytidine is illustrated by an observed substitution of the proline at position 223 in HSV-1 by isoleucine in equine herpesviruses type 1 (EHV-1) TK. Replacement of this proline residue, which has the capability of fostering a turn in the backbone conformation, with an isoleucine residue, which has a propensity to extend the backbone conformation maximally, appears to lead to the inability of the equine herpesviral TK to stabilize deoxycytidine as a substrate to be phosphorylated (Allen et al., 1979). This difference may be rationalized inasmuch as the range of orientations accessible to the side-chain of Glu-225 that forms a hydrogen bond to the substrate would be significantly different upon substitution of Pro-223 by isoleucine. If sequence variations at the specific residue positions seen in the crystal structure to be nearest neighbours to the thymidine are considered (Fig. 2), the EHV-1 TK sequence shares with VZV TK an identical pattern of differences to HSV-1 (His58Tyr, Met126Phe, Ala168Ser, Tyr172Phe). VZV TK, which retains a proline at position 223, actually seems to phosphorylate deoxycytidine at a faster rate than thymidine (Sawyer et al., 1986).

The part played by the loop segment (residues 215–226) thus relates to the ability of the protein to trap specific substrates. This observation is supported by the fact that an Asp215Arg substitution resulted in a lower catalytic rate with retention of apparent K_m (Michael et al., 1994). The increase in positive charge density within this loop region of the sequence (Fig. 2) would destabilize contacts of Arg-222 and Glu-225 with the substrate, which would decrease the ability of the protein to retain a properly oriented substrate. This mutational effect contrasts with that found for the ACV-resistant Arg176Gln mutant, Tr7, where the alteration at the substrate binding pocket changed the docking equilibrium K_m for both thymidine and deoxycytidine but did not markedly affect the catalytic rate (Larder et al., 1983). Taken together, these observations point to the coordination of specific residue segments in the catalytic process, with utilization of flexibility changes in the loop region to generate productive binding and the subsequent use of other ligands to provide target substrate specificity. This point is further exemplified by another of three reported HSV-1 mutants with resistance to drug BVdU, called B3, which was shown to result from the single substitution Ala168Thr; this variant shows wild-type behaviour with respect to thymidine and deoxycytidine, yet is resistant to ACV and BVdU (Darby et al., 1986).

The TK of HSV-2 and EHV-1, which have the substitution Ala168Ser in comparison to HSV-1 (Fig. 2), are also reported not to phosphorylate BVdU. The possibility exists that the more highly branched side-chain of serine or threonine either takes up too much space or alters the otherwise hydrophobic character of the binding pocket for the 5-substituent. The reported ease of accommodating five-membered heterocyclic substituents (De Winter & Herdewijn, 1996) suggests that the residue type has a more important effect than a modest increase in residue size in this context. While this line of reasoning fails to account for the ability of VZV TK to
phosphorylate BVdU, the effects of the further Leu169Thr substitution for VZV TK in comparison to HSV-1 TK could compensate for the effects of the Ala168Ser substitution given the flexible ligand associations characteristic of the substrate pocket.

Substrate binding and phosphorylation by the HSV-1 TK, in keeping with other kinases, are likely to involve domain closure leading to the formation of the transition state complex. Holding the nucleoside in place are contacts from the capping loop segment (residues 215–226) as well as contacts from residues located within a set of interacting helices. The helices which bound the nucleoside are linked to the rest of the structure by a network of distinctive interactions thereby facilitating the conformational shifts which presumably act to minimize the activation energy in going from substrate to product. The dynamic fit of the TK contact surfaces with the substrate therefore not only tolerates a range of ligand structure possibilities but also contributes to its range of drug-activating ability by enabling dynamic exchange of energy between the different substrates and the framework of the enzyme.

Recent work (Johansson & Karlsson, 1997) on the cloning of human cellular TK2 has led to the confirmation that TK2 and two of the other three known cellular deoxyribonucleoside kinases (deoxycytidine kinase and deoxyguanosine kinase, but not TK1) contain five out of the six regions of strong sequence homology previously identified for most herpesviral TK (Balasubramanian et al., 1990) and are also sequence-related to prokaryotic deoxyribonucleoside kinases. In particular, the cellular TK2 sequence contains nearly all of the specific residues seen to be closest to the bound nucleoside in the HSV-1 TK structure, i.e. Glu-83, Trp-88, Tyr-101, Gln-125, Arg-163, Tyr-172 and Glu-225 (HSV-1 TK numbering). It is also becoming evident that the narrow substrate specificity of cellular TK1 is rather unique since the third of the three known cellular deoxyribonucleoside kinases show broader nucleoside substrate specificities. In view of the implications for both the effectiveness and the toxicity of nucleoside analogues to be used in antiviral or anti-cancer therapies, the structure–function correlates for HSV-1 TK can serve as a guide for future research.

J.S.E. thanks Lawrence University (Appleton, WI 54912-0599, USA) for granting a sabbatical leave and the School of Biochemistry (Birmingham) for enabling participation in this work.

We wish to express our appreciation of support from Ian Skipper OBE and the Vaccine Research Foundation.

References


