Journal of Computer-Aided Molecular Design, 5 (1991) 385–404
ESCOM

J-CAMD 131

Computer-aided active-site-directed modeling of the Herpes Simplex Virus 1 and human thymidine kinase

Gerd Folkers[a,*], Susanne Trumpp-Kallmeyer[a], Oliver Gutbrod[a], Sabine Krickl[a], Jürgen Fetzer[b] and Günther M. Keil[b]

[a]*Pharmaceutical Institute of the University of Tübingen, Auf der Morgenstelle 8, 7400 Tübingen, Germany*
[b]*Federal Research Centre for Virus Diseases, Paul-Ehrlich Straße, 7400 Tübingen, Germany*

Received 23 March 1990
Accepted 8 March 1991

Key words: Molecular modeling; Herpes Simplex Virus 1; Thymidine kinase; Human thymidine kinase; Active sites; Interaction complexes

SUMMARY

Thymidine kinase (TK), which is induced by Herpes Simplex Virus 1 (HSV1), plays a key role in the antiviral activity of guanine derivatives such as aciclovir (ACV). In contrast, ACV shows only low affinity to the corresponding host cell enzyme. In order to define the differences in substrate binding of the two enzymes on molecular level, models for the three-dimensional (3-D) structures of the active sites of HSV1-TK and human TK were developed. The reconstruction of the active sites started from primary and secondary structure analysis of various kinases. The results were validated to homologous enzymes with known 3-D structures. The models predict that both enzymes consist of a central core β-sheet structure, connected by loops and α-helices very similar to the overall structure of other nucleotide binding enzymes. The phosphate binding site is made up of a highly conserved glycine-rich loop at the N-terminus of the proteins and a conserved region at the C-terminus. The thymidine recognition site was found about 100 amino acids downstream from the phosphate binding loop. The differing substrate specificity of human and HSV1-TK can be explained by amino-acid substitutions in the homologous regions.

To achieve a better understanding of the structure of the active site and how the thymidine kinase proteins interact with their substrates, the corresponding complexes of thymidine and dihydroxypropoxyguanine (DHPG) with HSV1 and human TK were built. For the docking of the guanine derivative, the X-ray structure of Elongation Factor Tu (EF-Tu), co-crystallized with guanosine diphosphate, was taken as reference. Fitting of thymidine into the active sites was done with respect to similar interactions found in thymidylate kinase. To complement the analysis of the 3-D structures of the two kinases and the substrate enzyme interactions, site-directed mutagenesis of the thymidine recognition site of HSV1-TK has been undertaken, changing Asp^{162} in the thymidine recognition site into Asn. First investigations reveal that the enzymatic activity of the mutant protein is destroyed.

*To whom correspondence should be addressed at Department of Pharmacy, ETH Zentrum, CH-8092 Zürich, Switzerland.

INTRODUCTION

Aciclovir (ACV), an acyclic guanine analog, is an effective inhibitor of Herpes Simplex Virus replication [1]. In order to express its antiviral activity, ACV requires phosphorylation by a viral thymidine kinase. In HSV-infected cells, ACV is phosphorylated to the monophosphate, which is subsequently converted to the di- and triphosphate by cellular kinases. The incorporation of the triphosphate form of ACV into the viral DNA results in the formation of a complex, which shows high affinity to the viral DNA polymerase and thereby inactivates the enzyme [2–4] (Fig. 1). In uninfected cells, the phosphorylation of ACV occurs only to a very limited extent, because of the low affinity of human TK to purine nucleosides. The lack of human TK to phosphorylate ACV and the reduced sensitivity of acyclo-GTP for α-DNA polymerase contributes to the extremely low cytotoxicity towards uninfected cells [5]. For the rational development of antiviral drugs, it was the goal of this study to determine the three-dimensional (3-D) structure of the host and the virus enzyme and the mode of binding that modulates the affinity of the substrates to the receptor site. If the 3-D structures of the active sites of the host and the virus enzymes are known, specific interactions of the substrates with the enzyme and the receptor-bound conformation can be determined. The knowledge of these substrate interactions could be extremely helpful in the design of new potent antiviral drugs.

METHODS

For the reconstruction of the active site of HSV1-TK and human TK, an integrated prediction of protein structure with simultaneous applications of primary, secondary and tertiary structure

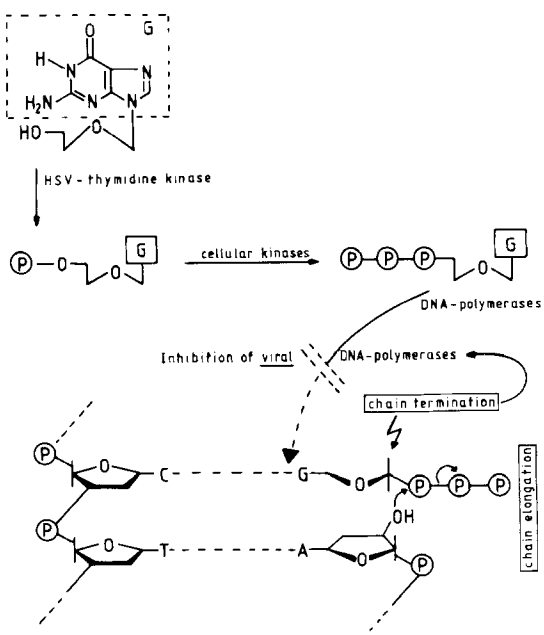


Fig. 1. Postulated mechanism of the ACV virostatic activity.

information was used [6]. The results obtained were validated to related isofunctional enzymes with known 3-D structure.

For the primary structure comparison, sequences of various TKs [7] were aligned by the method of Needleman and Wunsch [8]. To overcome several shortcomings of this method, visual inspection of mutant sequences of HSV-TKs with other TKs were carried out additionally, allowing for as much information as possible for each amino acid [9].

Secondary-structure prediction was performed by the method of Garnier et al. [10] using HYCON [11]. The nucleotide-binding enzymes adenylate kinase (ADK) [12] and elongation factor Tu (EF-Tu) [13] with known 3-D structures were taken as templates for the validation of the prediction and as reference molecules for the construction of the protein ligand complexes. In spite of their low percentage of homology, both enzymes show striking similarities in 3-D structures having parallel β-strands in their centers surrounded by α-helices.

The coordinates of ADK were taken from the Brookhaven Protein Data Bank [14]. The X-ray structure of EF-Tu was made available to us by J. Nyborg and T. Clark. Modeling was achieved with the molecular modeling package Sybyl 5.3. The interactive modeling and display of the structures was realized on an E&S PS390 graphics system, connected to a MicroVax 2 and a Vax 3500 via Ethernet/Decnet. The enzyme models were minimized and analysed with the AMBER force field [15], which is installed on a Convex C220. Additionally, interaction energies were obtained quantum-chemically in a supermolecule approach.

The procedure for site-directed mutagenesis will be published elsewhere.

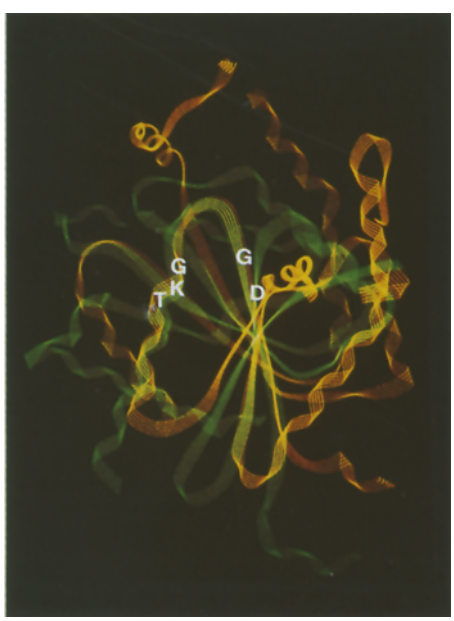


Fig. 2. Fit of the two nucleotide binding enzymes ADK and EF-Tu. The hydrophobic core of the two enzymes with the central twisted β-strands and the phosphate binding loop shows high correspondence. The conserved residues within the two enzymes are indicated in one-letter code.

RESULTS AND DISCUSSION

Proteins with functional homologies

The fact that the TKs belong to the family of kinases and to the class of nucleotide-binding enzymes allows structural comparison with the corresponding class of isofunctional proteins. We chose two proteins which had functional homologies in binding nucleotides in general, and especially guanosine derivatives, to account for the ACV affinity towards HSV1-TK.

Adenylate kinase, a phosphate-transferring enzyme, the structure of which has been defined by X-ray analysis [12], has been used as a template. It catalyses the reversible monophosphorylation from adenosine monophosphate to the diphosphate. Just as for the TKs, the mechanism of phosphate transfer is 'in line', following an SN_2 mechanism [16]. Investigations with transition-state inhibitors showed that both ADKs and TKs are inhibited by a number of tri-, tetra- and pentaphosphate dinucleotides [17]. Primary structures of ADK are known from a number of species [18]. Three-dimensional structures are described from ADK of porcine [19], rabbit [20] and yeast [21]. As a second nucleotide-binding enzyme, the crystal structure of EF-Tu, which has recently been solved by means of X-ray crystallography, was included in our investigations. Elongation factor Tu interacts with guanosine di- and triphosphate, which act as allosteric effectors, to control protein conformation required in the different steps of the elongation cycle [22]. The folding pattern of the protein around the diphosphate moiety is found to be very similar to that observed in other nucleotide-binding enzymes, and the residues involved in binding of GDP are highly conserved (Fig. 2).

Structural homologies and analogies within the group of kinases and phosphotransferases

Natural selection has produced families of proteins in which the amino-acid sequences have diverged widely, whereas the 3-D structures and biochemical functions have been conserved.

A commonly found feature of structural homology in nucleotide-binding enzymes is the so-called Rossman fold, consisting of two adjacent βαβ-units [23]. At least one βαβ-unit is required for the binding of one nucleotide.

A further characteristic of this type of enzyme is the phosphate binding site. It is localized in a special loop which connects the first β-strand of a parallel pleated sheet and an amphiphilic α-helix running antiparallel to this strand. The phosphate groups are fixed to the backbone of the loop by hydrogen bonds [24]. Additionally, this loop shows the highest degree of amino-acid conservation within all sequences.

The second important piece of structure information comes from the observation of Hélène and Lancelot [25] that guanine is the only base which interacts strongly with an E(D) R sequence. This is of special interest in this case, because most of the antiviral drugs against Herpes infections are guanosine derivatives, as mentioned in the Introduction. Interaction of guanine with this sequence disrupts the salt bridge E(D)···R, and guanine forms a complex with the carboxylate group by two hydrogen bonds. The guanine/carboxylate association is highly specific, and could play an important role in protein nucleic-acid recognition.

The above-mentioned data reveal that, first, a specific overall 3-D geometry is necessary for constructing a nucleotide binding site and, second, certain amino-acid side chains have to be associated with it, to confer substrate specificity.

Modeling of the HSV1-TK and human TK active site from primary and secondary structure analysis

The first step in modeling was to align sequences of several thymidine kinases with the sequences of the nucleotide binding enzymes ADK and EF-Tu to determine conserved regions which might be essential for substrate binding and conversion.

Three homologous regions are found within the thymidine kinases [26] (Fig. 3, regions A,B,C), two of these are also observed in ADK (regions A and C) [27]. Only one homologous region can be found in EF-Tu (region A). The HSV-TKs showed an additional homologous region to ADK (region D). The first one of these regions (A) corresponds to the nucleotide-binding site, and

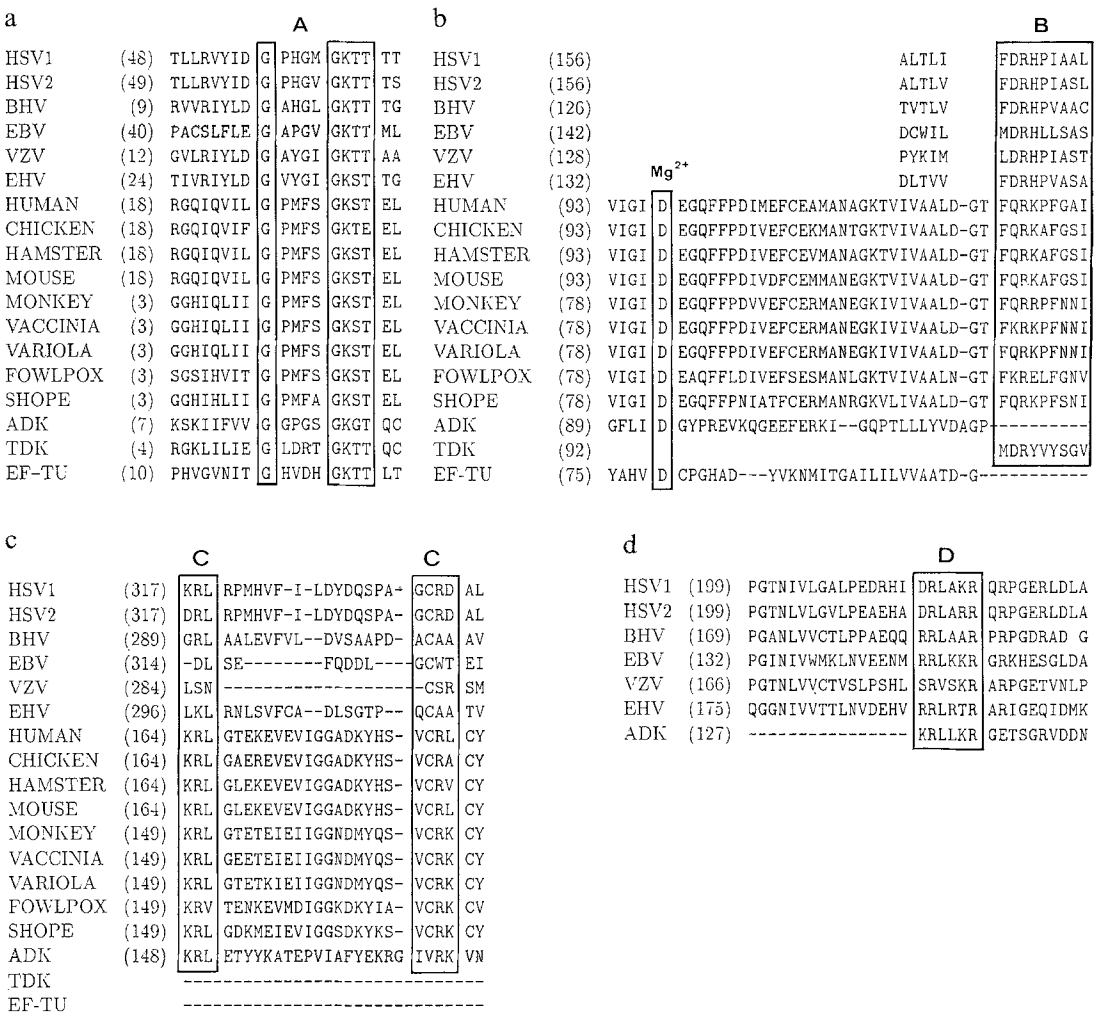


Fig. 3. (a) Primary structure homologies within the class of nucleotide binding enzymes. N-terminal region A with the phosphate binding loop. Conserved regions are framed. (HSV2, Herpes Simplex Virus 2 TK; BHV, Bovine Herpes Virus TK; EBV, Epstein-Barr Virus TK; VZV, Varicella Zoster Virus TK; EHV, Equine Herpes Virus TK; Human TK; Chicken TK; Hamster TK; Mouse TK; Monkeypox TK; Fowlpox TK; Shope, Shope Fibroma TK; ADK, Adenylate Kinase; EF-Tu, Elongation Factor Tu. (b) Homologous region B with the thymidine recognition site. (c) C-terminus with the homologous region C. (d) Additional basic region D of the HSV-TKs to ADK, not found in eukaryotic and pox TKs.

contains the flexible phosphate-binding loop with the highly conserved sequence G–X–X–X–X–G–K.

One striking feature is the occurrence of the amino acids Asp or Glu at the beginning of the loop in the Herpes viral thymidine kinases and in thymidylate kinase (TDK). In contrast to the other thymidine kinases, these enzymes are able to phosphorylate thymidine monophosphate to the diphosphate. The conservation of a negatively charged residue at that position might play a (yet unknown) role in thymidylate kinase activity.

The second homologous region (B) is located about 100 amino acids downstream from the phosphate-binding loop. The sequence F D R H in HSV-TKs and F Q R K in eukaryotic and pox TKs seems to be specific, at least for thymidine kinases. A very similar F E R K partial sequence can be found in A D K, which might shift the alignment shown in Fig. 3b for A D K to an alternative location of the gap, 17 residues earlier in the N-terminal direction. However, this would not be in agreement with the highly conserved region F C E R M, which shows up at this position for the mammalian and pox virus sequences.

Evidence for the importance of the F (X) R K sequence is given by the fact that neighbouring residues decide upon specific binding of Bromovinyldesoxyuridine (BVDU). The latter has no affinity to HSV2-TK, nor to a mutant of HSV1-TK (described by Larder and co-workers [38]) which bears a Thr at position 168 instead of an Ala [28]. The fact that the insensitive HSV2-TK shows a Ser at the same position reveals the importance of this region for substrate binding. However, only HSV-TKs bind acyclic guanine derivatives with considerable affinity. This leads to the hypothesis that the substitution of Gln for Asp in HSV-TKs might explain the specific binding of acyclic guanine derivatives to HSV-TKs. It is known, from guanine-binding enzymes, that side chains of the amino acids Asp and Gln or Arg define the specificity for guanine nucleotides, and that the exchange or lack of Asp normally results in a loss of binding affinity [29]. A very similar composition of amino acids is found in the F D R H sequence of the HSV-TKs.

The third conserved region (C) common to TKs and ADK is located at the C-terminus, and contains basic amino acids such as Arg and Lys, which could play a role in phosphate binding and transfer. Exceptions are both the BHV, which is known to be insensitive against aciclovir, and the EHV. Both sequences show the conserved cysteine residue, but followed by alanines instead of the basic amino acids found in the other sequences. ADK is the third exception, in having a valine at the conserved cysteine position. Up to now we have not been able to assign functionality to the exchanged residues. However, in ADK, the equivalent sequence is located just above the phosphate-binding loop, which might be a hint of its importance.

The additional homologous sequence (region D) of the HSV-TKs to ADK is located between the thymidine recognition site and the C-terminal conserved region. It consists of a number of basic amino acids, and might therefore play an essential role in substrate conversion. It can be assumed that this sequence is (as in ADK) placed in the neighbourhood of the phosphate-binding loop. This homologous region is absent in eukaryotic and pox TKs.

Compared to eukaryotic and pox TKs, the herpes viral enzymes are considerably larger. One insertion is located at the N-terminus of the proteins and is not necessary for the enzyme action. This could be shown in detail for a larger N-terminal region, represented by the first 45 amino acids [30]. Two additional insertions are placed between the thymidine recognition site and the C-terminal homologous region. Besides some conserved regions with amino acids, which preferably occur in turn regions, these insertions show high amino-acid exchanges within the sequences. The

conserved sequence parts may have an important directing influence in the folding mechanism of the polypeptide chain.

Although the primary structure analysis yielded only two small homologous regions for TKs and ADK, the secondary-structure prediction led to a pattern of alternating β-strands and α-helices with apparent homologies to ADK (Schemes 1–6).

In ADK, as well as in the TKs investigated, an amphiphilic helix, which follows the phosphate-binding loop, could be detected by comparison with the 3-D templates. This helix had been predicted, by the automated prediction algorithm, to be a β-sheet.

The folding of the human TK shows clearly the same secondary structure succession as found for ADK, with the magnesium-binding Asp^{97} on the third β-strand and the thymidine recognition site with the sequence F Q R K in a loop region after the fourth β-strand.

No commonly conserved aspartic acids for Mg binding are found within herpes viral TKs, which could give a hint of the location of the third β-strand. Instead of this, the HSV-TKs contain a conserved Asp or Glu at the beginning of the glycine-rich loop in the direct neighbourhood of Lys, which is involved in phosphate binding. This Asp or Glu might be involved in the binding of the Mg ion. This information was subsequently used for the location of the F D R H sequence in the HSV-TKs.

It is evident that the folding patterns of the TKs clearly represent a classical Rossman fold with typical βαβ sequence. Based on these results, models of HSV1-TK and human TK have been con-

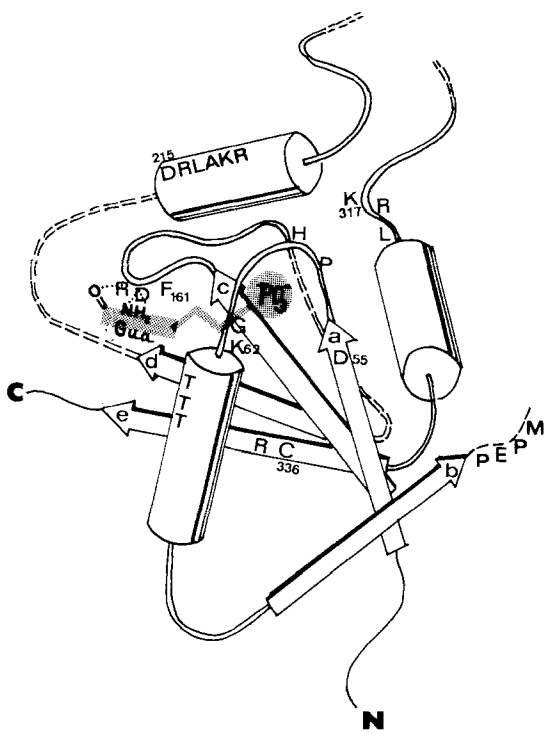


Fig. 4. Schematic representation of the predicted model of HSV1-TK substrate binding site with the bound substrate ACV. Conserved amino acids are indicated in one-letter code.

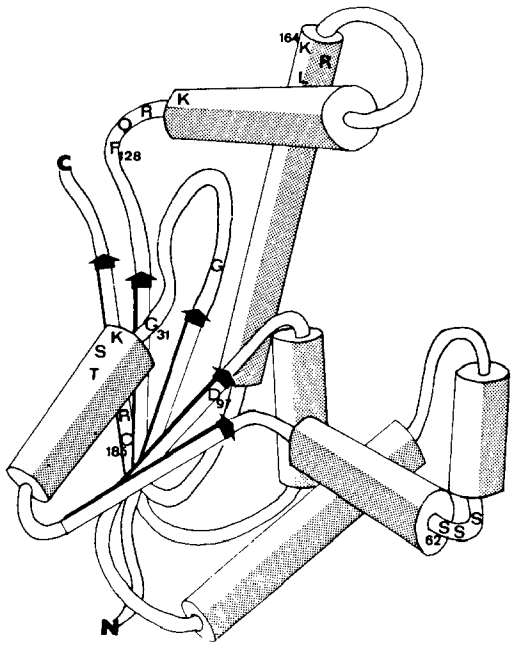


Fig. 5. Schematic representation of the human TK substrate binding site.

structed using the X-ray templates of ADK and EF-Tu, as are described below. A schematic representation of the two enzymes is given in Figs. 4 and 5.

The model of the active site of HSV1-TK consists of the first β-strand with the phosphate-binding loop and the amphiphilic helix. The second part of the molecule contains the fourth β-strand, with the F D R H sequence in a subsequent loop region. The third part includes an α-helix with the basic amino acids K R L at the N-terminus, and a turn connecting the last β-strand. The additional homologous region has a helical conformation, and is located above the phosphate-binding loop similar to ADK.

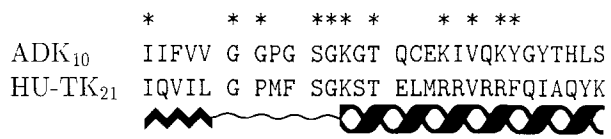
The model of the human TK shows high similarity to ADK, with five β-strands in the core region, connected by α-helices and loops. The amphiphilic helix, which follows the phosphate-binding loop, is highly conserved. It is thought that the α-helix generates an electrical field with a positively charged amino terminus. This positive field may have the effect of increasing the rate of nucleotide binding by positioning the phosphoryl groups favourably for catalytic hydrolysis [31]. The Mg-binding Asp^{97} is located on the third β-strand, in analogy to ADK. The succession of the β-strands places the thymidine recognition site with the sequence F Q R K in the neighbourhood of the phosphate-binding site in a loop region after the fourth β-strand. The third homologous region with the K R L sequence is situated in analogy to ADK above the nucleotide-binding site in a helical region.

3-D construction of the HSV1-TK

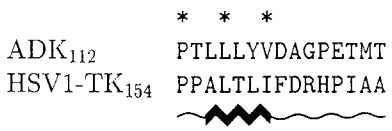
The 3-D structure of the HSV1-TK active site was built up using the X-ray structure of ADK as a template. In the case of an identical amino-acid sequence, the side chains of the ADK model were replaced with the homologous side chains of HSV1-TK. This replacement was done by re-

taining, as far as possible, the torsional angles common to the native and exchanged residues. In the case of insertions and deletions, the Cα-atoms of the homologous parts were superimposed, with the best possible fit, to the region of ADK. Several less-homologous parts of the enzyme, which form loops or α-helical regions and lie at the surface of the enzyme, could not be constructed, because they differ in length, in comparison to ADK, due to insertions or deletions.

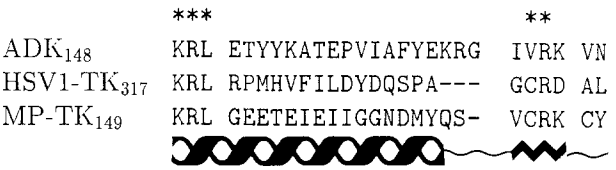
The construction of the first part of the active site contains the amino-acid sequence from amino acids 50–72 with the first β-strand, the phosphate-binding loop and the amphiphilic helix (Scheme 1)*. The second essential binding region contains the F^{161} D^{162} R^{163} sequence. The 3-D



arrangement of this part was determined from the 3-D arrangement of the β-strands in human TK (Scheme 2). The third essential binding region consists of the helix with the K^{317} R^{318} L^{319} se-



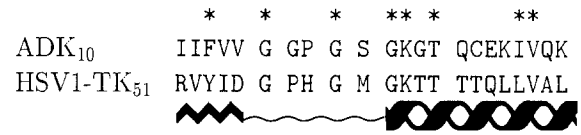
quence and the last β-strand. The most significant structural difference is a deletion of three residues in the helical region. For the construction, the K R L sequence was fitted onto the homologous counterpart in ADK (Scheme 3).



3-D construction of human TK

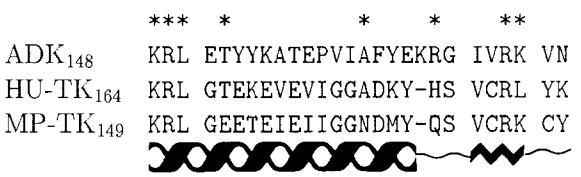
The 3-D structure of the human TK was built up in analogy to HSV1-TK and to ADK. The construction of the first part of the human TK started from amino acids 21–49. It includes the first β-strand with the phosphate-binding loop, the amphiphilic helix, turn and the second β-strand (Scheme 4). The construction of the following large helical region from 49 to 92 failed because of the low primary structure homology to ADK. Additionally, it shows high amino-acid exchange within the eukaryotic and ortho pox TKs. Therefore, it can be assumed that this region is located at the surface of the protein and does not play an essential role in substrate binding and conversion.

*As for the subsequent Schemes, secondary structures are indicated by zigzag lines (β-strand) and spiral turns (helices); positions of amino-acid similarities are shown by asterisks.



The Mg-binding Asp^{97} serves as a fixpoint for the location of the third β-strand. The fourth β-strand was constructed with respect to greatest homologies to ADK with the second homologous region F^{128} Q^{129} R^{130} K^{131} located in a loop region after the fourth β-strand. The helix which connects the third and the fourth β-strand shows two insertions at the C-terminal end, and had to be reconstructed (Scheme 5).

The last part of the human TK model consists of amino acids 164–192, and includes the third homologous region with helix, turn and the last β-strand. The helical region shows one deletion, compared with ADK. The last β-strand was aligned from additional homologies of the pox TKs to ADK. In analogy to ADK, they contain an Arg-Lys (RK) sequence (Scheme 6).



Geometry optimization

Energy minimization was performed with the molecular mechanics program AMBER, version 3.0 [32]. The minimization was done with the all-atom approach treating all hydrogens explicitly in the minimization procedure. To simulate solvent effects, a distance-dependent dielectric constant combined with a residue-based cutoff of 9 Å was used [33]. More realistic simulations for the solvent are not efficient in partly constructed models, because of the lack of large surface regions.

To achieve the removal of steric interactions and artefacts of the model-building with minimal change in the protein structure, positional constraints equal to 200 kcal/mol/$Å^2$ were applied to the first and the last amino acid in the secondary structure elements.

In Figs. 6 and 7, the constructed models of HSV1-TK and human TK are shown after geometry optimization. The highly conserved residues are indicated in one-letter code. Binding and conversion of the substrates have to be associated with these amino acids, which are shortly characterized as follows:

The glycine-rich loop with the sequence G–X–X–X–X–G–K shows the highest conservation within the nucleotide-binding enzymes, and is deeply buried in the catalytic cleft. From X-ray structures of different ADKs, it is evident that the loop binds one of the phosphate groups of the substrates, but the correct one among the four candidates has not been identified yet [6].

The amino acids F D (Q) R include the sequence D R, characteristic of the binding of guanosine nucleotides. It is located in the direct neighbourhood of the phosphate-binding loop in a coil region, following the fourth β-strand.

The conserved region K R L can be found both in ADKs and TKs. It is situated above the phosphate-binding loop, and might play a role in phosphate binding and transfer. The sequence C R can be found in nearly all thymidine kinases at the C-terminal end of the proteins. It is located behind the phosphate-binding loop on the last β-strand. (Darby and co-workers [38] showed that this sequence plays an essential role in substrate binding.) In an ACV-resistant mutant of HSV1-TK with reduced affinity to thymidine and bromovinyldesoxyuridine (BVDU), a single amino-acid substitution from Cys^{336} to Tyr^{38} is responsible for the altered enzyme activity.

The assumed Mg-binding Asp^{97} of the human TK is located on the third β-strand in topologically the same position as in ADK. No conserved Asp can be assigned in the HSV-TKs. Instead of this an Asp (Glu) at the beginning of the phosphate-binding loop might adopt the role of Mg-binding. It is located in the direct neighbourhood of the phosphate-binding Lys.

As described by Novotny et al. [34], criteria such as surface charge density, atomic packing, Coulomb-formula-based electrostatic interactions and the potential energy of misfolded and correctly folded structures are very similar, and therefore cannot be used for diagnosing incorrect folds. In order to estimate the quality of the modeled proteins, we therefore used the following additional criteria:

– the φ ψ angle distribution of the protein backbone;

– the intra- and interstructural hydrogen bond pattern of the secondary structure elements; and

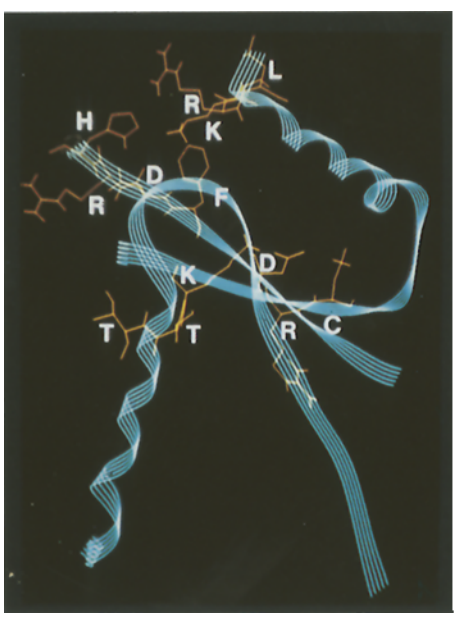


Fig. 6. Ribbon model of the geometry-optimized substrate binding site of HSV1-TK. The conserved amino acids which may take part in the substrate binding are indicated in one-letter code.

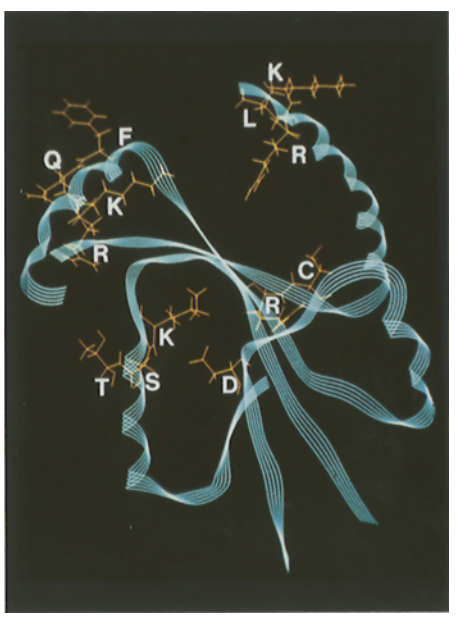


Fig. 7. Ribbon model of the geometry-optimized substrate binding site of human TK. The conserved amino acids which may take part in the substrate binding site are indicated in one-letter code.

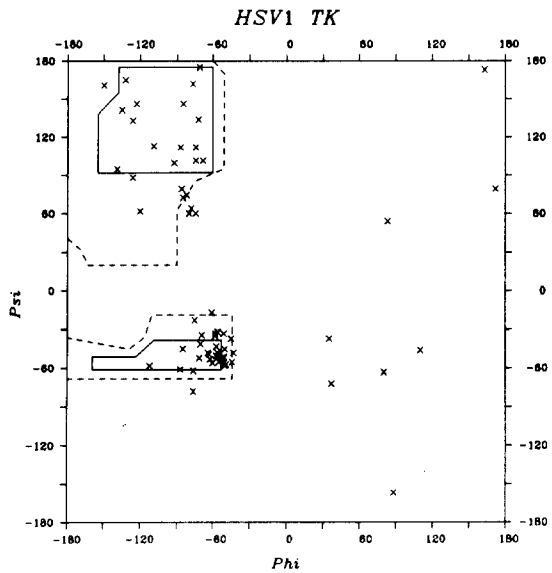


Fig. 8. Ramachandran diagram of HSV1-TK active site.

– the distribution of charged amino acids in the core region and at the outside of the modeled protein.

With the stiff peptide bond, and with rather rigid bond lengths, the polypeptide chain flexibility is restricted. In addition, the steric hindrance of the side chains prevents free rotation of the φ/ψ torsion angles. The conformational space of the peptide bond was investigated by Ramachandran et al. [35], and gives the highest density near $(-60°, -60°)$ for the right-handed α-helical conformation. The second maximum is near $(-90°, +120°)$, corresponding to β-strand conformation.

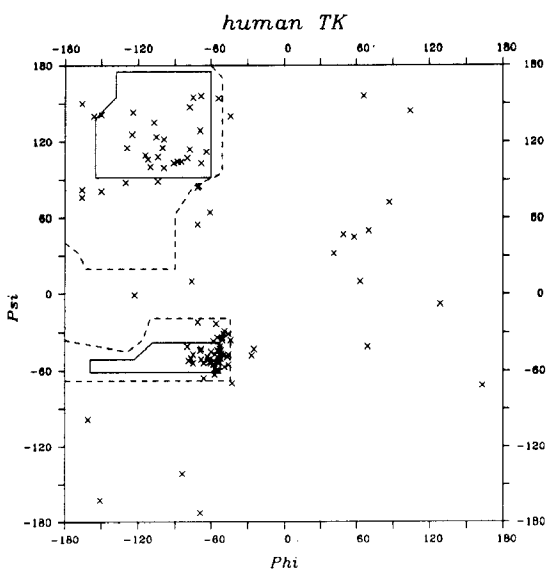


Fig. 9. Ramachandran diagram of human TK active site.

The analysis of the φ/ψ angles of our minimized models shows that about 10% of the entries lie in regions forbidden for all residues except glycine, a value which can be observed, in general, in other proteins (Figs. 8,9). For the HSV1-TK and the human TK these data points could be identified to be glycines as well.

Hydrogen bonds in the core region of a protein have important stabilizing effects. Burying the polar peptide backbone into the interior without loss of free energy is achieved only if hydrogen bonds within the peptide backbone can be formed. A survey of protein structures showed that about 90% of all internal polar groups form hydrogen bonds. This is consistent with the large quantity of defined secondary structure observed in proteins, and indicates that almost all possible hydrogen bond energy is scooped in.

In α-helices, the N–H—O angles are about 157°. The O–H distance average is 2.05 Å. The mean value for the hydrogen-bond distances and angles in β-strands are 1.96 ± 0.16 Å and $160° \pm 10°$, respectively [36].

In order to identify hydrogen bonds and salt bridges within our models, an interatomic distance criterion of 3 Å was used. Besides fairly good hydrogen-bond parameters in the modeled β-strands, the human TK as well as the HSV1-TK showed some bad values in the newly constructed α-helical regions. After 1500 steps of energy minimization (convergency reached), the hydrogen bonds in these parts were refined to mean average values found in other proteins. This geometry optimization yielded 155 main-chain and side-chain hydrogen bonds for the human TK active site (96 residues) and 110 hydrogen bonds for the HSV1-TK active site (70 residues).

From protein modeling it is known that incorrectly modeled structures have more buried residues than is normal. Rashin and Honig [37] found that charged groups are only rarely buried in native proteins and, if so, that they are always involved in hydrogen bonds or salt bridges. To estimate the quality of the modeled proteins, the number of buried ionizable groups may therefore indicate whether the models are correctly folded.

As expected, the majority of the charged side chains of human and HSV1-TK were located at the surface of the molecules exposed to the solvent. Exceptions are the phosphate-binding lysines (Lys^{67} in HSV1, Lys^{32} in human TK) in the glycine-rich loop and the Mg-binding Asp^{97} in human TK on the third β-strand. These residues can be found in a similar 3-D arrangement in ADK.

The refined structures, with the hydrophobic amino acids colored in orange and the charged amino acids type-coded, are shown in Figs. 10 and 11, respectively. Except for the amino acids mentioned above, no charged residues are located in the core region.

Substrate interaction homologies in adenylate kinases, dehydrogenases and elongation factor Tu

Besides the same overall 3-D geometry of nucleotide binding enzymes the above-named have geometrically similar substrate binding sites with highly conserved interactions.

In porcine ADK, the binding sites of the substrates ATP and AMP are found at geometrically similar positions relative to the β-sheet as they are in the NAD-binding site of the dehydrogenases [39]. In addition, the sheet topology is similar, that is, the position of ATP in porcine ADK corresponds exactly to that of the adenosine-containing moiety of NAD in lactate dehydrogenase (LDH).

The adenine moiety of ATP is probably H-bonded to the phenolic OH group of Tyr^{95} in ADK; the corresponding amino acid in LDH is Tyr^{85}. The location of AMP in ADK corresponds exactly to the position of nicotinamide in LDH.

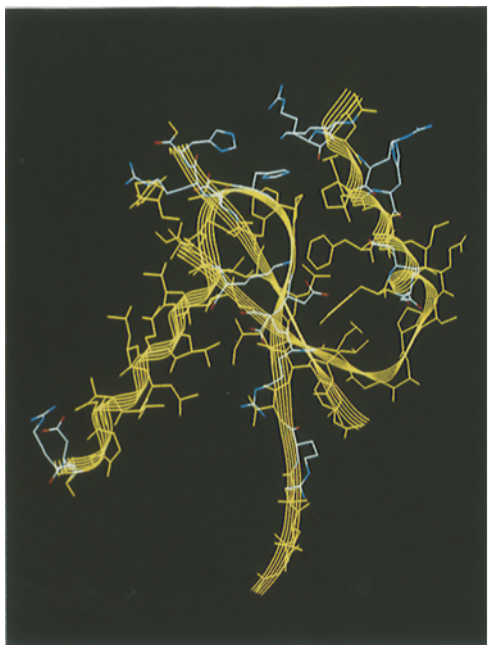


Fig. 10. Minimized model of the HSV1-TK active site.

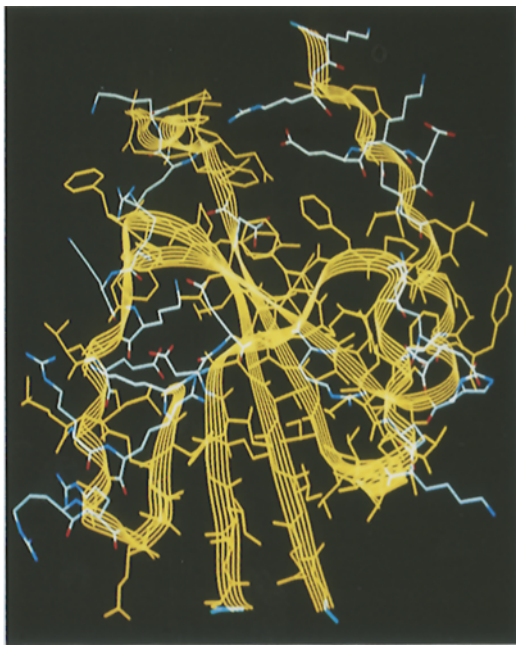


Fig. 11. Minimized model of the human TK active site, with the hydrophobic amino acids colored orange and the charged amino acids coded by atom type.

In contrast to the ATP situation, no special interactions of AMP with the ADK protein could be observed. To define the exact binding mode of AMP, numerous attempts have been made with magnetic resonance experiments and chemical modification studies [20,40]. However, a convincing detailed proposal of the correct binding site is lacking so far. The phosphate-binding site also shows striking similarities between dehydrogenases and adenylate kinases in containing a flexible loop region, which is able to undergo large conformational changes during nucleotide binding [41].

The binding region of EF-Tu has the same secondary structure arrangement as found in ADKs, and the substrate GDP is located in geometrically the same position as one adenosine moiety of the transition-state inhibitor AP_5A in yeast adenylate kinase (X-ray structures are shown in Figs. 12 and 13, respectively) [42]. In contrast to adenine in ADK, the guanine exhibits specific interactions with the enzyme. In EF-Tu, GDP is positioned at the carboxy end of the central β-strands 1, 4 and 5 with the guanine moiety somewhat exposed to the solvent. The high specificity towards guanine nucleotides in EF-Tu originates from the formation of hydrogen bonds of the amidino function of guanine to the charged carboxyl side chain of Asp^{43}. The ribose ring faces the protein with its endo side, leaving the 2′ and 3′ hydroxyl groups exposed to the solvent. The binding site for the diphosphate residue of GDP is made up by the glycine-rich loop which is commonly found in nucleotide-binding enzymes. The interaction of the α-phosphate appears to be only low, whereas the binding of the β-phosphate is much stronger, additionally involving an Mg ion which is

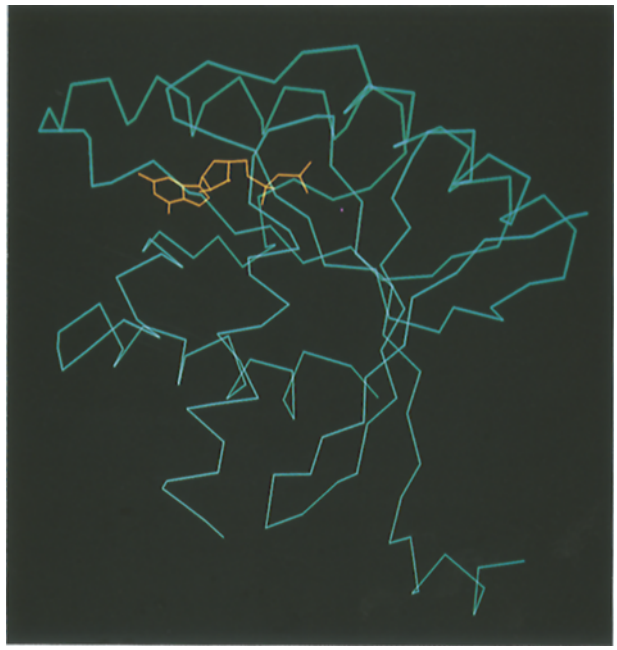


Fig. 12. C_α-atom backbone of Elongation Factor Tu from *E. coli* with bound GDP.

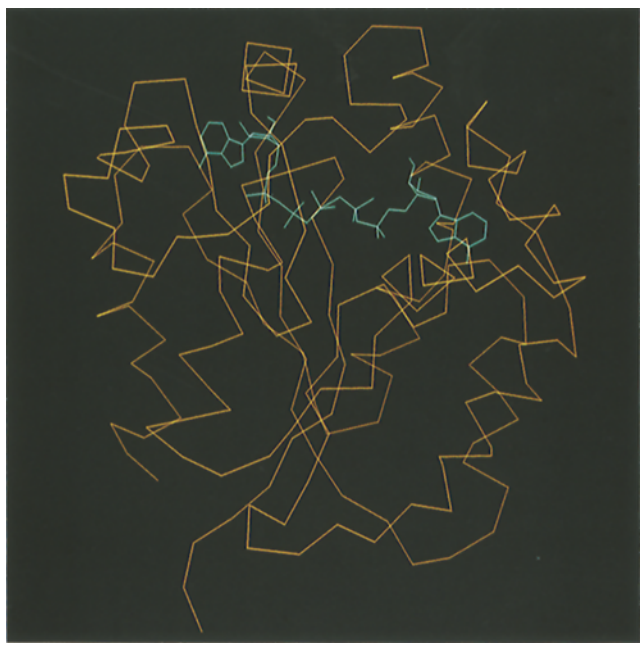


Fig. 13. C_α-atom backbone of yeast adenylate kinase with the bound inhibitor AP_5A. The AMP binding site is assumed to be located on the left side above the nucleoside binding loop.

bound to Asp^{80}. A fit of the secondary-structure elements of the two enzymes reveals that the Mg-binding aspartic acids are located in the same position.

The binding region of AMP in porcine ADK, as well as the binding region of GDP in EF-Tu, corresponds to the putative binding region with the thymidine recognition site in human and HSV1-TK. In the following steps these analogies will be used for the construction of the enzyme substrate interaction complexes.

Incorporation of the substrates into the active sites of HSV1 and human TK

For the construction of the interaction complexes of the guanine derivative DHPG-monophosphate with HSV1-TK, EF-Tu was taken as the reference molecule. The docking of the substrate into the active site was achieved with respect to the following conditions:

– For the specific binding of the guanine base, the characteristic amino-acid sequence D R is necessary [25]. This sequence can be found in the homologous F D R H region, located in a loop at the carboxy end of the fourth β-strand in a similar position to EF-Tu.

– The phosphate is located in the phosphate-binding loop, known from several X-ray structures of nucleotide-binding enzymes.

– The stereochemical configuration of the open-chain sugar moiety (pro-chiral) has to be that of desoxy-ribose [44].

The specific affinity of guanosine derivatives to HSV-TKs might be explained by the fact that, in eukaryotic and pox TKs, the acidic amino acid Asp in F D R H is changed into the neutral Gln, which might restrict the complex formation with the guanine base. In order to prove that assumption, Asp^{162} of HSV1-TK was changed into Asn in a site-directed mutagenesis study. The exchange results in a mutant protein with destroyed enzymatic activity [45].

For the construction of the complex, DHPG-MP was fitted into the active site of HSV1-TK in analogy to GDP in EF-Tu. The guanine moiety of DHPG forms three hydrogen bonds between the amino and amid nitrogen of the guanine ring to Asp^{162}, and from the carbonyl function of the base to Arg^{163}. The acyclic sugar of DHPG is positioned above the phosphate-binding loop with O5′ in identical position with respect to O5′ of GDP in EF-Tu.

Thymidine was fitted into the active sites of the two enzymes in a similar position, by using additional structural information from thymidylate synthase. In 1986, the atomic structure of thymidylate synthase (TS) from *Lactobacillus casei* was determined at 3 Å resolution [46]. The enzyme catalyses the conversion from dUMP to dTMP. The native enzyme consists of a dimer of identical subunits with a central antiparallel five-stranded β-sheet connected by α-helices. The dUMP (dTMP) is bound within a cavity which is lined by the side chains of 25 residues. The primary sites of substrate binding involve the conserved amino-acid residues Tyr^{217}, Gln^{218}, Arg^{219}, Arg^{179} and Asn^{229}. From model-building, it can be assumed that Arg^{219} is directly involved in the binding of dUMP (dTMP), building-up electrostatic interactions with one of the electron-rich carbonyl oxygens of the pyrimidone ring or with the phosphate group of the substrate.

Compared with thymidine kinases, the conserved sequence Y Q R in TS shows high chemical and geometrical correspondence to the highly conserved sequence F Q R in thymidine kinases. Therefore, we constructed a complex of HSV1 and human TK with thymidine by building hydrogen bonds and/or electrostatic interactions from O4 of the pyrimidine ring to the Arg residue in the thymidine-kinase-specific sequence. The sugar moiety was placed in analogy to the sugar moiety of GDP in EF-Tu, with the O3′ exposed to the solvent.

TABLE I
RMS DEVIATION (Å) OF THE MINIMIZED INTERACTION COMPLEXES

HSV1-TK + thymidine	0.161
HSV1-TK + DHPG	0.189
Human TK + thymidine	0.168

Geometry optimization of the interaction complexes

Geometry optimization of the interaction complexes was performed with the same methods as described before to exclude bad contacts and steric interactions after model-building. Substrate charges were taken from MNDO calculations. Although we are aware of the dynamic behavior of proteins, our first goal was the construction of static models. They describe the position of the substrates in the active site within a certain arrangement of specific amino acids. For the description of protein/ligand interactions, dynamic simulations of the whole complexes will be performed.

The final structures were compared with the starting structures using a root-mean-square (r.m.s.) procedure. The results are given in Table 1. Figures 14 and 15 show the optimized interaction complexes of human and HSV1-TK with thymidine. The active site of HSV1-TK with DHPG is shown in Fig. 16.

A more detailed investigation of the substrate interaction was done with respect to the mutagenesis experiments. In a supermolecule approach [47], DHPG was surrounded by a guanidinium cation and an acetic acid anion, representing arginine and aspartic acid, respectively, in a geometry similar to that of the active site. Geometry optimization was performed semiempirically using MOPAC and having the DHPG moiety fixed.

In a second experiment, we chose the same starting geometry, but exchanged the acetic acid for the corresponding amide, which represented the asparagine residue. The geometry optimization of the complex yielded a different interaction geometry for the energy minimum, as expected, accompanied by a considerable loss in total energy. Although this approach could be criticized because of the Hamiltonians used, in our opinion we get clues for subsequent docking procedures.

CONCLUSION

The modeling of the active sites, or parts thereof, of HSV1-TK and human TK, respectively, by a knowledge-based protein prediction procedure, led for the viral and cellular thymidine kinases to an overall active-site architecture very similar to all other nucleoside-binding proteins. The model geometry is able to explain most of the experimental results known up to now, including our own mutagenesis experiment. This clearly assigns considerable importance for the reaction mechanism to the Asp^{162} which had been located in the active site by the computer modelling*. Although the model represents only a minor part of the whole protein, we might be able to decide upon structural alterations of substrate analogues by dynamic docking experiments, and use this information for QSAR studies. This procedure is currently being investigated.

*The model coordinates are available upon request from G. Folkers in mol or pdb file formats.

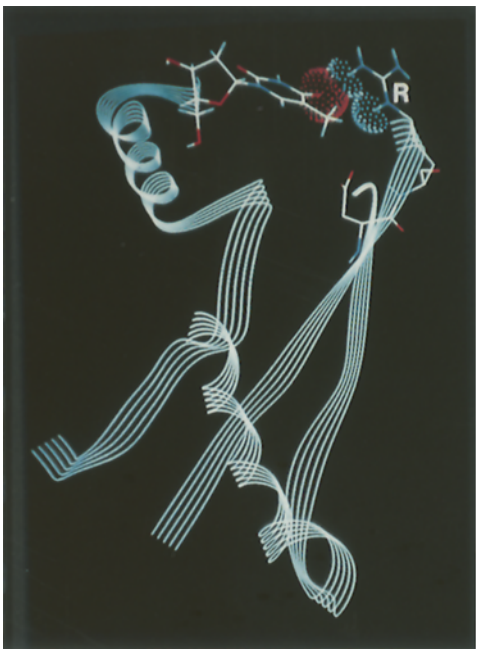


Fig. 14. Minimized interaction complex of HSV1-TK with thymidine, hydrogen bonded to Arg^{163}.

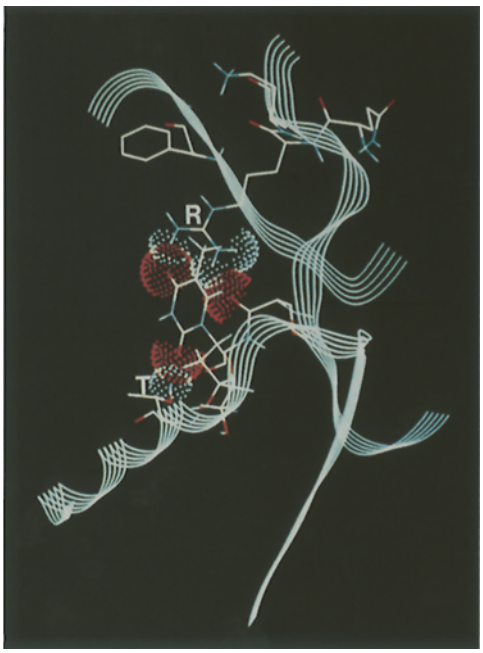


Fig. 15. Minimized interaction complex of human TK with thymidine which forms hydrogen bonds from O4 of the base to Arg^{130} and from O2 to Thr^{34}.

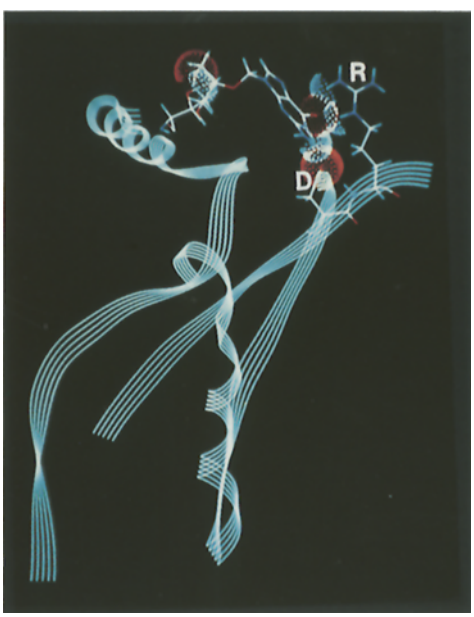


Fig. 16. Minimized interaction complex of HSV1-TK with DHPG. Guanine forms hydrogen bonds from the amino and amide function to Asp^{162} and from the carbonyl function to Arg^{163}. An additional hydrogen bond is formed from O3′ of the acyclic sugar to Thr^{64}.

Considerable support for the docking of the nucleosides into HSV1-TK comes from the fact that the relative spatial orientation of DHPG and thymidine is very similar to that found in an earlier study by active analogue approach [44], the 5-methyl-group of thymidine pointing in the direction of the N1 in the DHPG six-membered ring (Figs. 14 and 16). This orientation is in full agreement with the experimental observation that substituents in position 5 of thymidine strongly affect binding affinity and specificity.

In summary, it seems that this combined theoretical and experimental approach gives enough information to establish models, which can now serve as starting points for more detailed studies.

ACKNOWLEDGEMENT

This work has been supported by the BMFT, grant no. 0318750A. The authors are indebted to Prof. W. Schwöbel for helpful discussions, and to one of the referees for his useful remarks on the alignment problems.

REFERENCES

1 Elion, G.B., Furmann, P.A., Lubbers, C.M., De Miranda, P., Beauchamp, L. and Schaeffer, H.J., Proc. Natl. Acad. Sci. USA, 74 (1977) 5716.
2 Fyfe, J.A., Keller, P.M., Furmann, P.A. and Elion, G.B., J. Biol. Chem., 253 (1978) 8721.
3 Elion, G.B., Am. J. Med., 73 (1982) 7.
4 Reid, R., Mar, E.C., Huang, E.-S. and Topal, M.D., J. Mol. Biol., 263 (1988) 3898.
5 Furmann, P.A., St. Clair, M.H. and Spector, T., J. Biol. Chem., 259 (1984) 9575.
6 Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E. and Thornton, J.M., Nature, 326 (1987) 347.
7 a. Bradshaw, H.D. and Deininger, P.L., Mol. Cell. Biol., 4 (1984) 2316.
b. Hofbaur, R., Wien, personal communication.
c. Kwoh, T.J. and Engler, J.A., Nucl. Acids Res., 12 (1984) 3959.
d. Lewis, J.A., Mol. Cell. Biol., 6 (1986) 1998.
e. Boyle, B.D., Coupar, B.E.H., Gibbs, A.J., Seigman, L. and Both, G.W., Virology, 156 (1987) 355.
f. Upton, C. and McFadden, G., J. Virol., 60 (1986) 920.
g. Otsuka, H. and Kit, S., Virology, 135 (1984) 316.
h. Robertson, G.R. and Whalley, J.M., Nucl. Acids Res., 16 (1989) 11303.
i. Keil, G., BFA Tübingen, personal communication.
j. Davison, A.J. and Scott, J.E., J. Gen. Virol., 67 (1986) 1759.
k. Wagner, M.J., Sharp, J.A. and Summers, W.C., Proc. Natl. Acad. Sci. USA, 78 (1981) 1441.
8 Needleman, S.B. and Wunsch, C.D., J. Mol. Biol., 48 (1970) 443.
9 Folkers, G., Krickl, S. and Trumpp, S., Arch. Pharm. (Weinheim), 322 (1989) 409.
10. a. Garnier, J., Osguthorpe, D.J. and Robson, B., J. Mol. Biol., 120 (1978) 97.
b. Levin, J.M., Robson, B. and Garnier, J., FEBS Lett., 205 (1986) 303.
11 Troeger, W. and Jung, G., unpublished (1987).
12 Schulz, G.E., Elzinga, M., Marx, F. and Schirmer, R.H., Nature, 250 (1974) 120.
13 La Cour, T.F.M., Nyborg, J., Thirup, S. and Clark, F.C., EMBO J., 4 (1985) 2385.
14 Allen, F.H., Bellhard, S.H., Brice, M.D., Cartwright, G.A., Doubleway, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.A. and Watson, D.G., Acta Crystallogr., B 35 (1979) 2331.
15 Weiner, P.K. and Kollman, P.A., J. Comp. Chem., 2 (1981) 287.
16 Bone, R., Cheng, Y.-C. and Wolfenden, R., J. Biol. Chem., 261 (1986) 5731.
17 Arnold, J.R.P., Cheng, M.S., Cullis, P.M. and Lowe, G., J. Biol. Chem., 261 (1985) 1985.
18 Schulz, G.E., Schiltz, E., Tomasselli, A.G., Frank, R., Brune, M., Wittinghofer, A. and Schirmer, R.A., Eur. J. Biochem., 161 (1986) 127.

19 Shyy, Y.-J., Tian, G. and Tsai, M.-D., Biochemistry, 26 (1987) 6411.
20 Fry, D.C., Kuby, S.A. and Mildvan, A.S., Biochemistry, 24 (1985) 4680.
21 Tomasselli, A.G. and Noda, L.H., Eur. J. Biochem., 132 (1983) 109.
22 Barbacid, M., Annu. Rev. Biochem., 56 (1987) 770.
23 Rao, S.T. and Rossman, M.G., J. Mol. Biol., 76 (1973) 241.
24 Schulz, G.E. and Schirmer, R.H., Principles of Protein Structure, Springer, Heidelberg, 1979.
25 Hélène, C. and Lancelot, G., Progr. Biophys. Mol. Biol., 39 (1982) 1.
26 Folkers, G., Trumpp, S., Eger, K., Roth, H.J. and Schwöbel, W., Med. Sci. Res., 15 (1987) 633.
27 Folkers, G. and Trumpp, S., Med. Sci. Res., 15 (1987) 1495.
28 Dever, T.E., Glynias, M.J. and Merrick, W.C., Proc. Natl. Acad. Sci. USA, 84 (1987) 1814.
29 Field, H.J., Darby, G. and Wildy, P., J. Gen. Virol., 49 (1980) 115.
30 Marsden, H.S., Haarr, L. and Preston, C.M., J. Virol., 46 (1983) 434.
31 Hol, W.G., Progr. Biophys. Mol. Biol., 45 (1985) 149.
32 Kollman, P., Weiner, S.J., Case, D.A., Singh, U.C., Ghio, D., Alagona, G., Profeta, S. and Weiner, P., J. Am. Chem Soc., 106 (1984) 765.
33 Karplus, M. and McCammon, A., Annu. Rev. Biochem., 52 (1983) 263.
34 Novotny, J., Rashin, A.A. and Bruccoleri, R.E., Proteins, 4 (1988) 19.
35 Ramachandran, G.N. and Sasisekharan, V., Adv. Prot. Chem., 23 (1968) 283.
36 Baker, E.N. and Hubbard, R.E., Progr. Biophys. Mol. Biol., 44 (1984) 97.
37 Rashin, A.A. and Honig, B.H., J. Mol. Biol., 173 (1984) 503.
38 Larder, B.A., Derse, D., Cheng, Y.C. and Darby, G., J. Biol. Chem., 258 (1983) 2027.
39 Pai, E.F., Sachsenheimer, W. and Schirmer, R.H., J. Mol. Biol., 114 (1977) 37.
40 Ray, B.D., Rösch, P. and Rao, B.D.N., Biochemistry, 27 (1988) 8669.
41 Dreusike, D. and Schulz, G.E., FEBS Lett., 208 (1987) 301.
42 Egner, U., Tomasselli, A.G. and Schulz, G.E., J. Mol. Biol., 195 (1987) 649.
43 Jurnak, F., Science, 230 (1985) 32.
44 Folkers, G., Sakahara, K., Schwöbel, W. and Eger, K., Arch. Pharm. (Weinheim), 322 (1989) 395.
45 Fetzer, J., Ph.D. Thesis, University of Tübingen (1991).
46 Hardy, L.W., Finer-Moore, J.S., Montfort, W.K., Jones, M.O., Santi, D.V. and Stroud, R.M., Science, 25 (1987) 448.
47 Loew, G., Nienow, J.R. and Poulsen, M., Mol. Pharmacol., 26 (1984) 19.