Structural Basis for Recognition of Polyglutamyl Folates by Thymidylate Synthase[†,‡]

Alexander Kamb,[§] Janet Finer-Moore,[§] A. Hilary Calvert,[∥] and Robert M. Stroud*[,§]

Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, and Drug Development Section, Royal Marsden Hospital, Sutton, Surrey, England

Received March 27, 1992; Revised Manuscript Received July 17, 1992

ABSTRACT: Thymidylate synthase (TS) catalyzes the final step in the de novo synthesis of thymidine. In vivo TS binds a polyglutamyl cofactor, polyglutamyl methylenetetrahydrofolate (CH_2-H_4folate), which serves as a carbon donor. Glutamate residues on the cofactor contribute as much as 3.7 kcal to the interaction between the cofactor, substrate, and enzyme. Because many ligand/receptor interactions appear to be driven largely by hydrophobic forces, it is surprising that the addition of hydrophilic, soluble groups such as glutamates increases the affinity of the cofactor for TS. The structure of a polyglutamyl cofactor analog bound in ternary complex with deoxyuridine monophosphate (dUMP) and *Escherichia coli* TS reveals how the polyglutamyl moiety is positioned in TS and accounts in a qualitative way for the binding contributions of the different individual glutamate residues. The polyglutamyl moiety is not rigidly fixed by its interaction with the protein except for the first glutamate residue nearest the *p*-aminobenzoic acid ring of folate. Each additional glutamate is progressively more disordered than the previous one in the chain. The position of the second and third glutamate residues on the protein surface suggests that the polyglutamyl binding site could be utilized by a new family of inhibitors that might fill the binding area more effectively than polyglutamate.

TS[1] catalyzes formation of TMP from dUMP. Because of its essential role in DNA synthesis, TS is an attractive target for anticancer drugs, including the widely used chemotherapeutic 5-fluorouracil, and it is the focus of much current research and development in the pharmaceutical industry (Harrap et al., 1989; Appelt et al., 1991). All organisms express TS, and TS is perhaps the most highly conserved enzyme throughout phylogeny. It is a functional dimer composed of two identical subunits that together form two active sites. The TS reaction mechanism involves transfer of a methylene group from a cofactor, CH_2-H_4folate, onto the 5 position of the dUMP pyrimidine ring, followed by hydride transfer (Pogolotti et al., 1986; Finer-Moore et al., 1990). Upon binding substrate and cofactor, the protein undergoes a conformational change that orients the ligands and compacts the protein (Montfort et al., 1990).

The cofactor used in vivo includes a string of glutamate residues attached to the benzoyl moiety of the cofactor via an amide bond. Folate containing a single glutamate is imported into cells through a specific transport system (Nahas et al., 1972; Henderson, 1990). Inside the cell, folate serves as a substrate for a glutamyl transferase enzyme that sequentially attaches glutamate residues using γ-linkages. The end result is a bulk population of folates that contains predominantly 4–7 attached glutamyl residues per folate molecule. Thus, from bacteria to mammals, all enzymes that use folates as substrates or cofactors recognize polyglutamyl folates. Yet, in vitro all enzymes that require folates recognize the monoglutamyl molecule as well.

Three hypotheses have been proposed to explain the role of folate polyglutamylation (Schirch & Strong, 1989). First, the additional glutamate residues could help trap folates inside the cell by increasing folate solubility (Hoffbrand et al., 1973). Second, the glutamyl moiety could assist in the transfer of folates from one enzyme to another; for example, CH_2-H_4-folate once oxidized by TS could be delivered efficiently to dihydrofolate reductase for regeneration via an enzyme channeling mechanism (Paquin et al., 1985). Third, different numbers of glutamyl residues could confer different affinities for the various folate-binding enzymes, thereby allowing the cell to regulate how much folate is utilized in a particular pathway (Lewis et al., 1979).

Despite determination of a large number of crystal structures of TS and other folate-binding proteins (Hardy et al., 1986; Oefner et al., 1988; Montfort et al., 1990; Perry et al., 1990; Matthews et al., 1990; Davies et al., 1990; Bystroff & Kraut, 1991), nothing is known about the detailed structural basis for recognition of polyglutamyl folates by enzymes. Here we describe the crystal structure of *Escherichia coli* TS bound in complex with substrate and a polyglutamyl cofactor analog, CB3717polyglu (Figure 1).[2] The complex not only reveals the structural consequences of the interaction between the polyglutamyl moiety and TS but it also reveals how polyglutamylation enhances cofactor binding to TS. In addition, the structure suggests some alternative routes for future drug design efforts.

[†] Supported by National Institutes of Health Grants RO1-CA-41323 (to J.F.-M. and R.M.S.) and GM24485 (to R.M.S.). Crystallographic refinement was carried out at the Pittsburgh Supercomputer Center under Grant DMB890040P.

[‡] The coordinates have been deposited in a Brookhaven Protein Data Bank File under 1BBQ.

* Author to whom correspondence should be addressed.

[§] University of California.

[∥] Royal Marsden Hospital.

[1] Abbreviations: CH_2-H_4folate, methylenetetrahydrofolate; TS, thymidylate synthase; dUMP, deoxyuridylate; PABA, *p*-aminobenzoic acid; TMP, thymidylate; CB3717, 10-propargyl-5,8-dideazafolate; CB3717polyglu, polyglutamyl CB3717; EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol; glu-*n*, *n*th γ-linked glutamate residue of CB3717polyglu; 5-FU, 5-fluorouracil.

[2] By convention we use the *L. casei* amino acid numbering scheme when referring to amino acid residues by number. For example, for *E. coli* residues $1 \geq n \leq 89$, n (*L. casei*) = n (*E. coli*) + 2; for residues $n > 89$, n (*L. casei*) = n (*E. coli*) + 52.

0006-2960/92/0431-9883$03.00/0

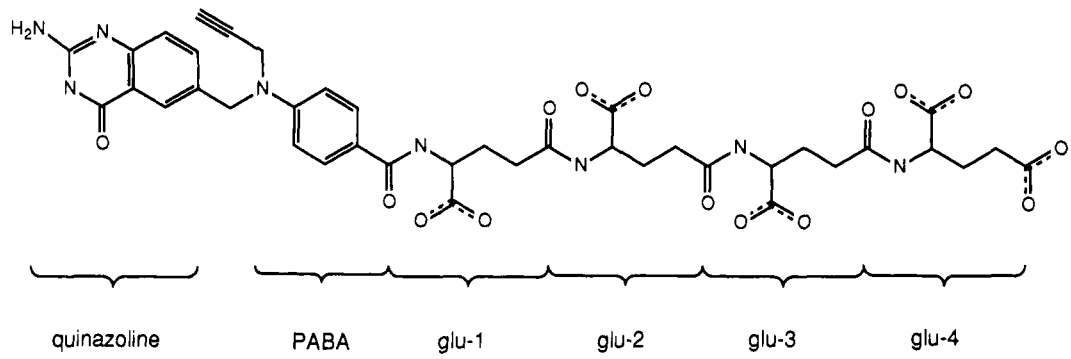

FIGURE 1: Polyglutamyl folate cofactor analog CB3717polyglu. The major structural groups are labeled with the exception of the triple-bonded propargyl moiety.

EXPERIMENTAL PROCEDURES

Reagents. Purified *E. coli* TS was supplied by Drs. F. and G. Maley (Belfort et al., 1983). dUMP was purchased from Sigma. CB3717polyglu was obtained as a gift from Dr. H. Calvert and Imperial Chemical Industries (ICI). The compound was analyzed by mass spectrometry and shown to have a molecular mass of 847 Da, the mass predicted for CB3717 containing four γ-linked glutamate residues (data not shown).

Crystallization. Crystallization was as described previously (Montfort et al., 1990). TS was dialyzed against 20 mM potassium phosphate (pH 8.0), 0.1 mM EDTA, and 2 mM DTT. A protein/ligand solution was prepared with TS (5.3 mg/mL), 100 mM dUMP, 2 mM CB3717polyglu, and 10 mM DTT and incubated overnight at 4 °C. Crystals were grown by vapor diffusion in hanging drops after the protein/ligand solution was mixed in different ratios ranging from 1:4 to 4:1 with well buffer consisting of 20 mM potassium phosphate (pH 8.0), 2.3 M ammonium sulfate, and 10 mM DTT to achieve a final volume of 10 μL per drop. Crystals grew over a 2-week interval at room temperature. Prior to capillary mounting for X-ray crystallography, crystals were transferred to depression wells containing 20 mM potassium phosphate (pH 8.0), 2.5 M ammonium sulfate, 0.1 mM EDTA, and 10 mM DTT and soaked overnight at room temperature.

X-ray Data Collection. Diffraction data were collected from a single crystal at room temperature using a Nicolet IPC area detector on a three-circle goniometer as described previously (Montfort et al., 1990). Cu Kα X-rays from a rotating anode X-ray generator (Rigaku) with Franks' mirrors were used. The crystal was a hexagonal rod roughly 700 μm long and 250 μm wide. The space group was $P6_3$ with unit cell dimensions $a = b = 127.3$ Å and $c = 68.2$ Å. Each data frame involved an oscillation of 0.2° collected for 180 s. Reflections were indexed and intensities were integrated using SCAN (Blum et al., 1987). Structure factors were scaled together using SCALEI and merged with XENGEN (Howard et al., 1985).

Structure Solution, Refinement, and Model Building. The crystal structure of the TS/dUMP/CB3717polyglu ternary complex was deduced from difference Fourier maps using phases calculated from the TS/dUMP/CB3717 ternary complex of Montfort et al. (1990) with the ligands removed. The initial $|F_o| - |F_c|$ and $2|F_o| - |F_c|$ density maps were ambiguous in the region of the polyglutamyl moiety. The structure, omitting the polyglutamyl moiety, was refined using X-PLOR (Brunger, 1990) energy minimization (150 cycles) followed by atomic *B*-factor refinement (40 cycles). The $2|F_o| - |F_c|$ electron density maps were calculated and used to build in the polyglutamyl moiety (Table I). For map calculation, data was 50.0 to 2.3 Å with a 2σ cutoff were used. A bulk solvent mask was used to improve the quality of the low-resolution phases (>5.0 Å). The model was then subjected to 1.35 pS of molecular dynamics refinement involving a slow cool from 3000 to 300 K using X-PLOR with all charges turned off.

Electrostatic Calculations. Electrostatic energy surfaces for the *E. coli* TS protein in the ternary complex conformation with the ligands removed and for the *Lactobacillus casei* protein in the open, unbound conformation were calculated using the DELPHI program which solves a linearized form of the Poisson–Boltzman equation (Honig et al., 1989). Histidines were given a net charge of +0.5 except for *E. coli* His-53 whose charge was set to 0. The solvent dielectric was set at 78.5 and the solute dielectric at 2.0. A full Coulombic potential was used with a grid spacing of 0.65 Å and an ionic strength of 145 mM.

Synthesis of Benzoyl Pentaglutamate. Pentaglutamate with two γ-linkages followed by two α-linkages was obtained from the Biomolecular Resource Center at UCSF. The peptide resin (0.1 mmol) was suspended in 6 mL of dimethylformamide (DMF) and mixed with 0.3 mL of triethylamine followed by 0.5 g of benzoic anhydride. After the solution was allowed to react for 3 h at room temperature, the resin was washed with 300 mL of DMF followed by 300 mL of chloroform. The resin was dried and the benzoylated peptide removed from the resin with hydrofluoric acid. The benzoyl pentaglutamate was subjected to mass spectrometry and shown to have a mass of either 768 or 790 Da, consistent with the predicted mass of 745 Da plus one or two sodium ions, respectively. TS inhibition assays were performed as described (Pogolotti et al., 1986).

RESULTS

Structure of the Polyglutamyl Moiety and Ligands. The 2.3-Å resolution electron density map clearly revealed the positions of the ligands including most of the polyglutamyl moiety. The first three residues (Glu-1, Glu-2, and Glu-3) of the string of four γ-linked glutamate residues could be positioned at the first active site within the electron density (Figure 2). However, the density for Glu-4 was ambiguous. Attempts to model Glu-4 into a region of nearby electron density followed by refinement yielded equivocal results; no connection was observed for this density at a contour level of 1σ in $2|F_o| - |F_c|$ difference Fourier maps. Thus, this density may derive either from Glu-4 or from an ordered water molecule. The density in the second active site for Glu-1 and Glu-2 was unambiguous, but no connected density for Glu-3 and Glu-4 was visible (data not shown). In both active sites there is a trend toward less sharply defined position from Glu-1 toward Glu-4. This is illustrated by the *B*-factors of

Table I: Statistics for Crystallographic Data and Crystallographic Refinement Statistics of TS/dUMP/CB3717polyglu

(a) Statistics for Crystallographic Data[a]

resolution range (Å)	no. of possible reflections	no. of collected reflections	% of possible reflections collected	no. of total observations	av $I/\sigma I$	R_{sym} (%)
infinity–4.07	5 161	4 090	79.2	15 877	25.64	5.41
4.07–3.23	5 048	5 005	99.1	14 852	12.95	8.40
3.23–2.82	5 033	4 946	98.3	12 148	5.18	16.46
2.82–2.56	5 045	4 804	95.2	9 527	2.80	24.36
2.56–2.38	5 011	4 218	84.2	6 825	1.82	30.81
2.38–2.24	4 996	2 344	46.9	3 092	1.31	34.94
sum: infinity–2.24	30 294	26 307	86.8	62 321	9.22	9.11

(b) Crystallographic Refinement Statistics[b]

av *B*-factors (Å²)			parameters		stereochemical ideality		diffraction agreement		rms differences between monomers	
	monomer	monomer 2								
waters	27		no. of atoms	4540	bonds (Å)	0.015	*R*-factor (%)	18.2	protein (all atoms) (Å)	0.956
protein	14	19	no. of waters	104	angles (deg)	3.3	no. of reflections	24 302	protein (Cα) (Å)	0.355
dUMP	11	13	no. of discretely	0	Ramachandran	0	resolution (Å)	50–2.3	dUMP (Å)	0.943
folate	10	14	disordered residues		violations		σ cutoff	0	CB3717 (Å)	0.818
Glu-1	32	29								
Glu-2	60	46								
Glu-3	70									

[a] $R_{sym} = [\sum_{hkl}\sum_{i=1}^{N} w_i(I_{av} - I_i)^2 / \sum_{hkl}\sum_{i=1}^{N} w_i(I_i)^2]^{1/2}$, where $I_{av} = 1/N\sum_{i=1}^{N} I_i$ and $w_i = 1/\sigma_i^2$. [b] *R*-factor $= \sum||F_o| - |F_c||/|F_o|$.

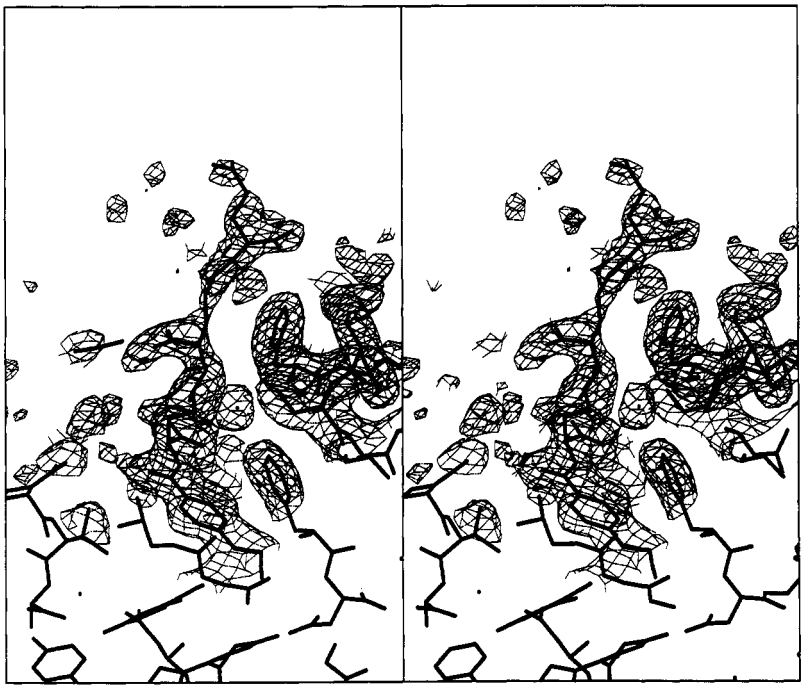

FIGURE 2: Divergent stereo depiction of electron density for the polyglutamyl moiety in active site 1. Only the first three γ-linked glutamate residues are included in the model. Also visible within the electron density depicted is the His-53 imidazole.

the atoms: In the first active site, the Cγ *B*-factors range from 32 Å² for Glu-1 to 74 Å² for Glu-3 (Table Ib). This trend likely reflects increasing vibrational motion of the polyglutamyl group as it progresses out from the cofactor core structure. Nonetheless, the model for the polyglutamyl moiety fits the experimental data well. The *R*-factor of the structure is low (18.2%), and the high resolution of the electron density map defines the orientation of peptide carbonyl and α-carboxylic acid groups for residues in the polyglutamyl moiety.

The model for the second active site was built without reference to the first active site. Despite the inability to locate Glu-3 in the second active site, the general configuration of the glutamate residues that could be seen is the same in both active site regions, as are the orientations of TS side chains that interact with the polyglutamyl moiety (Figure 3). The similarity of the two active site structures (Table Ib) is not forced by crystal symmetry because the unit cell contains a complete enzyme dimer per asymmetric unit. As in the crystal structure of the ternary complex with monoglutamyl CB3717, the *B*-factors for the second monomer are on average higher than those for the first monomer (19 vs 14 Å²), perhaps because of differences in crystal contacts. This increased disorder may prevent the highly mobile Glu-3 residue from contributing significantly to the electron density map in the second active site. To simplify discussion of the ligand conformation, the focus will be on the first active site except where explicitly stated.

The polyglutamyl group does not significantly affect positions of atoms in the cofactor analog and substrate relative to ligand atoms in the TS/CB3717/dUMP ternary structure (Montfort et al., 1990). Both ligand molecules in the polyglutamyl ternary complex are highly ordered, and the rms deviations in the positions of ligand atoms in this structure

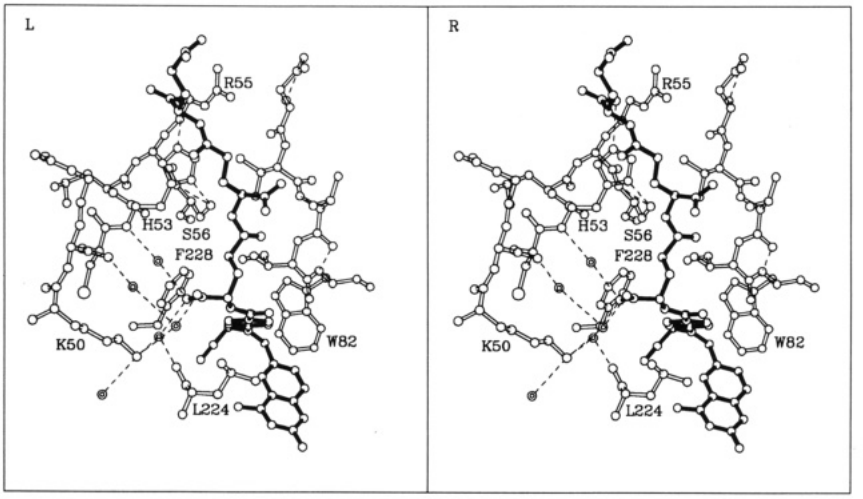

FIGURE 3: Divergent stereo drawing of the modeled polyglutamyl moiety in active site 1. CB3717polyglu is depicted with solid bars; pertinent side chains are depicted with open bars; hydrogen bonds are depicted with dashed lines; ordered water molecules are depicted with doughnuts.

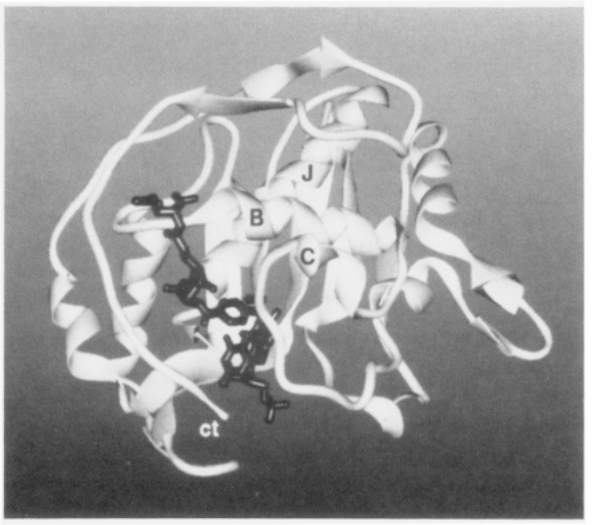

FIGURE 4: Ribbon diagram of TS monomer 1 with dUMP and CB3717polyglu bound. The ligands are in black. Helices B, C, and J are labeled, as is the C-terminus (ct).

compared to the positions of equivalent atoms in the monoglutamyl CB3717 ternary complex are small: 0.37 Å for dUMP and 0.33 Å for all atoms in CB3717 excluding the glutamates. The greatest effect of the polyglutamyl moiety is to decrease the temperature factors of the CB3717 atoms. Whereas the average *B*-factors of the protein (16 Å^2) and dUMP (11 Å^2) are the same in the polyglutamyl structure as in the monoglutamyl complex, the *B*-factors for CB3717 are lower. The average *B*-factor for the glutamate moiety (Glu-1 in the polyglutamyl CB3717) decreases from 48 Å^2 in the monoglutamyl structure to 32 Å^2. Therefore, the binding of the polyglutamyl moiety to TS does not reorient the ligands, but it does further restrict their motion inside the active site.

The polyglutamyl group lies exposed along the protein surface in an extended configuration (Figures 3 and 4). Nine of 12 dihedral angles along the polyglutamyl backbone are within 56° of the angle for a maximally extended chain, 180° (Table II). The average value of these nine torsion angles is 155°. Whereas Glu-1 is largely buried within the active site cavity with 11% of its surface accessible to solvent, more than half the surfaces of Glu-2 and Glu-3 are exposed to solvent (Table III). Glu-2 lies in the center of a short, broad canyon approximately 10 Å wide and 5 Å deep between the C helix and a segment of extended chain that leads to the C-terminus. Glu-2 does not approach the edges of this canyon. Glu-3 lies

Table II: Torsion Angles along the Polyglutamyl Moiety Backbone[a]

atoms				active site 1 (deg)	active site 2 (deg)
C1	NP1	CA1	CB1	63	80
NP1	CA1	CB1	CG1	59	57
CA1	CB1	CG1	CD1	−168	−165
CB1	CG1	CD1	NP2	−124	−119
CG1	CD1	NP2	CA2	149	158
CD1	NP2	CA2	CB2	−146	−145
NP2	CA2	CB2	CG2	−162	−130
CA2	CB2	CG2	CD2	164	−175
CB2	CG2	CD2	NP3	−145	83
CG2	CD2	NP3	CA3	178	
CD2	NP3	CA3	CB3	42	
NP3	CA3	CB3	CG3	157	

[a] CA is Cα, CB is Cβ, etc. The numbers correspond to the residue number to which the atom belongs.

Table III: Solvent Accessibility of Individual Residues in the Polyglutamyl Moiety either While Bound to Active Site 1 of TS or When Free in Solution[a]

residue	solvent-accessible area (Å^2) bound	free	bound/free (%)
Glu-1	19.2	172.2	11.2
Glu-2	87.9	161.2	54.5
Glu-3	154.2	235.8	65.4

[a] Solvent-accessibility values were calculated using the method of Connolly (1983) with a probe radius of 1.4 Å.

completely exposed on the surface in a region with little or no concave curvature. Thus the polyglutamyl moiety originates within the confines of the active site pocket and stretches out in an extended conformation along the surface of TS, exposed to solvent.

Conformation of the Protein. The conformation of the protein in the TS/dUMP/CB3717polyglu structure is nearly identical to the protein conformation found in the ternary complex with monoglutamyl CB3717 (Montfort et al., 1990). The rms deviation between backbone atoms (C, O, N, Cα) of the two ternary complexes is 0.30 Å. To estimate the standard deviation of the differences between these structures as a function of temperature factor, $\sigma_{xyz}(B)$, atoms were grouped according to *B*-factor and Gaussian distributions were fit to the histogram of atom positional shifts between structures, Δr, for different *B*-factor ranges. A function of the form

$$\sigma_{xyz}(B) = aB^2 + bB + c$$

(B = *B*-factor) was fit to a plot of the standard deviations of

each of the Gaussian distributions versus *B*-factor (Perry et al., 1990; Fauman, in preparation). The positions of only 77 atoms (1.8% of the total atoms in the dimer) differ by more than 3σ above the expected difference for the atoms based on their *B*-factors. All of these are side-chain atoms. The largest side-chain shifts are for His-53 in the B helix and Phe-228 in the highly conserved hydrophobic J helix which makes up part of the core of the protein (Perry et al., 1990). The His-53 imidazole rotates roughly 90° as if to form a floor for the polyglutamyl moiety that passes directly across it. Phe-228, which is conserved in 17 of 18 published TS sequences, forms part of the PABA ring binding site and is one of only two residues which undergoes side-chain conformational changes to accommodate ligands when the ternary complex forms (Montfort et al., 1990). In both active sites of the complex with CB3717polyglu, the side chain of Phe-228 rotates about 50° around the Cα–Cβ bond toward the polyglutamyl moiety. However, its phenyl ring remains perpendicular to and within van der Waals contact of the PABA ring. Because Phe-228 lies immediately below His-53, it is likely that the rotation of His-53 induces the motion of the Phe-228 side chain. Thus, the effect on protein structure caused by binding of the polyglutamyl group is limited to His-53 and its immediate environment.

Interactions between the Polyglutamyl Group and TS. The polyglutamyl group makes several contacts with the protein surface. The α-carboxylic acid of Glu-1 interacts through well-ordered water molecules (*B*-factors $\sim$20 Å^2) with Lys-50 and several main-chain atoms including His-53 (NH), Leu-224 (O), and Ile-310 (O) (Figure 3). The Lys-50 side chain is well-ordered. Its Nζ coordinates a hydrogen-bonding network which links several noncontiguous parts of the polypeptide chain including the C-terminus and helices A, B, J, and K. All the atoms that participate in hydrogen bonds with Lys-50 have *B*-factors less than 20 Å^2 in both the native and ternary complex structures. Glu-2 lies directly above the imidazole group of His-53. However, it does not make a salt bridge with His-53 (which presumably carries little positive charge due to its location amid several basic residues) because the α-carboxylic acid moiety points in the opposite direction from the imidazole, that is, into the solvent region. The plane of the amide between Glu-2 and Glu-3 is oriented roughly parallel to the imidazole ring of His-53, and Glu-2 does not form hydrogen bonds with the imidazole. Thus, His-53 contacts the glutamate moiety passively and simply rotates out of the way to avoid a steric clash. The only obvious attractive contact that Glu-2 is able to make is an electrostatic interaction with Lys-311, the Nϵ of which lies 5.77 Å from one of the α-carboxylic acid oxygens. The backbone carbonyl group of Glu-3 approaches the His-53 imidazole ring and the guanidinium group of Arg-55 closely. The modeled oxygen (Glu-3) is 3.0 Å from the His-53 Nϵ and 3.4 Å from the Arg-55 NH1. However, the uncertainty in positioning the atoms in this region of the structure due to their high *B*-factors prevents unambiguous identification of hydrogen bonds. In summary, with the exception of Glu-1, the polyglutamyl moiety forms no clearly discernible hydrogen bonds with TS.

Of the amino acid residues that interact with the polyglutamyl moiety, those that lie closer to Glu-3 have higher *B*-factors than those that lie nearer to Glu-1. The Nζ of Lys-50 has a *B*-factor of 13 Å^2; the imidazole nitrogens of His-53 have *B*-factors of 25 Å^2, and the guanidinium group of Arg-55 has atoms with *B*-factors of almost 60 Å^2. This suggests that the ensemble, polyglutamyl moiety residues and contacting amino acid atoms, may be rigidly positioned only near Glu-1. Beyond this anchor point, the entire ensemble may vibrate either in unison or in an independent manner.

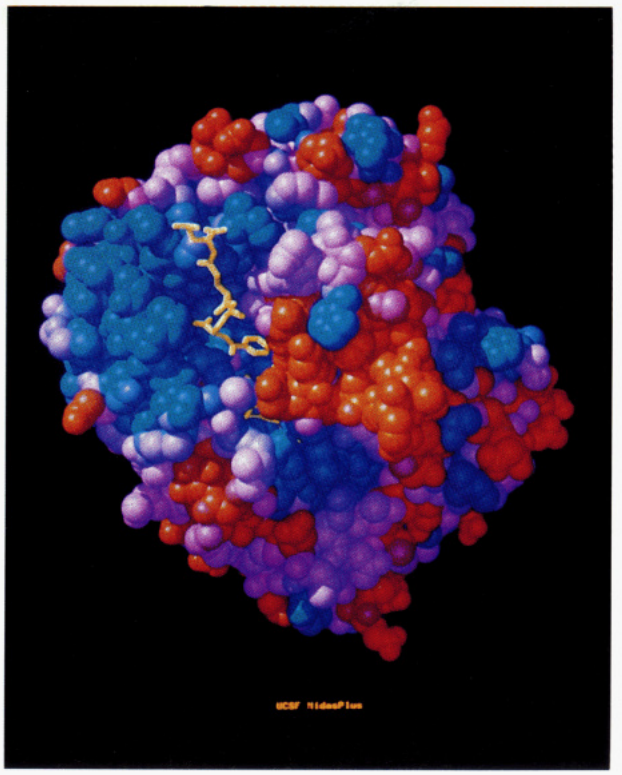

FIGURE 5: CPK model of the TS ternary complex with monoglutamylated CB3717 (Montfort et al., 1990). dUMP is omitted and CB3717polyglu is shown as a yellow stick model. The protein atoms are colored according to the value of the electrostatic potential at their centers. Blue represents values >+2.5 kcal; red represents values <–2.5 kcal; lavender areas are approximately neutral.

Calculation of electrostatic surface potentials for the *E. coli* ternary complex structure without the polyglutamyl moiety showed that the negatively charged glutamate residues lie in a position where they can interact favorably with surface regions of positive potential (Figure 5). Glu-1, Glu-2, and Glu-3 reside well within areas of positivity. Even though His-53 is uncharged, the surface of *E. coli* TS has a positive electrostatic potential (>+1.7 kcal) at all His-53 atom positions. Interestingly, the positive charge contours drop off immediately beyond Glu-3, suggesting that additional glutamate residues would not contribute much electrostatic binding energy to the *E. coli* complex. Calculation of electrostatic surface potentials for *L. casei* TS reveals a similar picture of the electrostatic environment in the homologous region of the protein. The surface is positively charged in the locale that would be in contact with the polyglutamyl moiety if it adopted a similar conformation in *L. casei* TS as in *E. coli* TS (data not shown). Thus, for both *E. coli* and *L. casei* TS, the polyglutamyl binding site on the protein presents an attractive, charged surface to the cofactor, suggesting that electrostatic interactions play a major role in the binding interaction between the polyglutamyl moiety and the TS molecule.

Polyglutamate without the folate is an extremely poor inhibitor of TS. A compound that contains a benzoyl group linked to five glutamate residues has an IC_{50} greater than 2 mM with *L. casei* TS (data not shown). For comparison,

Table IV: Relative Binding Affinities of Polyglutamyl Folate Molecules or Folate Analogs to either *E. coli* (Friedkin et al., 1975), *L. casei* (Matthews et al., 1987), Pig (Lu et al., 1984), or Human (Sikora et al., 1988) TS[a]

glutamate residue no.	*E. coli* TS	*L. casei* TS	pig TS	human TS
1	1	1	1	1
2	10	5	33	26
3	67	29	50	87
4	67	36	167	119
5	100		100	113
6	200	414	83	
7			67	

[a] Numbers are normalized to the monoglutamyl folate affinity value in each species; it is not valid to compare directly values from different columns of the table.

folate inhibitors with five glutamate residues have K_i values with pig or *E. coli* TS 100 times higher than monoglutamyl inhibitors (Table IV). This demonstrates that when polyglutamate is attached to a hydrophobic folate molecule, the additional glutamates contribute significantly to the binding interaction. However, even with the addition of a PABA ring mimic, polyglutamate is too hydrophilic and too flexible to bind TS well.

DISCUSSION

Polyglutamylation Alters the Kinetics of Ligand Binding. Polyglutamyl CH_2-H_4folates are better substrates for TS than monoglutamyl folate. For example, human TS displays a 2-fold higher V_{max} and nearly a 20-fold lower K_m with tetraglutamyl CH_2-H_4folate than with CH_2-H_4folate (Dolnick & Cheng, 1978). Because the positions of the ligands are identical in the ternary complex regardless of the polyglutamyl group, it is likely that the enzymatic preference for polyglutamyl cofactor form does not involve conformational differences in the catalytic sites. Rather, the improvement in catalytic behavior probably reflects enhanced binding conferred by the polyglutamyl moiety on the cofactor.

Several studies using γ-linked polyglutamyl folates have shown that polyglutamyl folates and folate analogs have a higher affinity for TS than simple folates but the overall binding energies contributed by the polyglutamyl moiety vary between species (Table IV). For *E. coli* TS, the K_i for folate antagonists is nearly 70-fold lower with tetraglutamyl folate antagonists than for monoglutamyl folates (Friedkin et al., 1975). For pig TS, the K_i drops 167-fold as the number of glutamate residues increases from one to four (Lu et al., 1984). For *L. casei*, the K_i falls about 36-fold (Matthews et al., 1987; Thorndike et al., 1990).

The fraction of binding energy contributed by each glutamate also appears to depend on the species of TS examined. In *E. coli*, the difference between monoglutamyl and diglutamyl inhibitors corresponds to a 10-fold decrease in K_i. The addition of a third glutamate causes a further 7-fold drop in K_i while a fourth glutamate has no significant effect. Addition of two more glutamates for a total of six leads to a 3-fold further decline. With pig TS, K_i values fall 33-fold in going from one glutamate to two glutamates, whereas the addition of a third glutamate decreases K_i less than 2-fold. A fourth glutamate lowers K_i by another factor of 2, but increasing the polyglutamyl moiety length beyond four residues actually interferes with binding to pig TS (Lu et al., 1984). Results reported for *L. casei* TS are different in certain respects. The addition of a second glutamate moiety decreases K_i 5-fold while a third glutamate causes a 6-fold further drop. The addition of a fourth glutamate has only a minimal effect, but incorporation of two more glutamates leads to a further 11-fold decline in K_i. Thus, with *L. casei* TS the first four glutamate residues contribute relatively less to binding than with pig and *E. coli* TS. However, the next two residues, Glu-5 and Glu-6, contribute much more binding energy in *L. casei* compared to pig and *E. coli*.

Structural Interpretation of Ligand Binding Experiments. It should be possible to account qualitatively for these differences in binding contribution of the various glutamate residues using structural information (Table IV). In *E. coli* TS, Glu-2 passes directly over His-53 and may interact electrostatically with Lys-311. In a diglutamyl inhibitor complex, the γ-carboxylic acid would presumably lie adjacent to Arg-55, capable of forming a strong electrostatic contact. Glu-3 lies well within a region of positive electrostatic potential and is expected to contribute further binding energy. The observation that Glu-4 does not improve cofactor binding for *E. coli* TS is consistent with the ternary structure. Glu-4 is not visible in the electron density map, suggesting that it is highly disordered and does not interact strongly with the protein surface.

Studies using polyglutamyl inhibitors show that the relative binding energies of various polyglutamyl inhibitors for pig TS are qualitatively similar to relative binding energies for human TS. The pig TS amino acid sequence has not been determined. However, given the extraordinary homology of TS sequences, it is probable that the pig sequence differs little from the human or mouse sequences which are identical to one another in the proposed polyglutamyl moiety binding region (Figure 6). If the polyglutamyl moiety adopts a similar structure with pig and human TS as with *E. coli* TS, Glu-2 would pass over the top of a phenylalanine residue that occupies the position equivalent to His-53. The site equivalent to Lys-311 is also occupied by lysine in human, mouse, and presumably pig TS, providing the possibility of electrostatic binding energy. The polyglutamyl moiety might then continue toward the position of Arg-55, which in the mammalian sequences is filled by lysine. Thus, the overall picture of bonding interactions with pig TS might be analogous to that in *E. coli*. Similar electrostatic interactions appear possible as well as the potential for more extensive hydrophobic contacts between the backbone segment of Glu-2 and the phenylalanine at the His-53 site. These hydrophobic forces may explain why the contribution of Glu-2 to the overall binding energy is greater in pig than in *E. coli*.

In *L. casei* TS, the situation is more complicated. The sequence is more divergent in this B-helix region of the protein that would be in contact with the polyglutamyl moiety if the structure of the ternary complex were similar. Nevertheless, the assumption that the polyglutamyl moiety adopts the same basic conformation in *L. casei* as in *E. coli* leads to a satisfactory understanding of the binding interactions. In *L. casei* the His-53 position is occupied by proline. It is possible that this prevents Glu-2 from making optimal contacts with the protein in this area possibly because of the conformational rigidity of proline. This may explain why Glu-2 is more important for binding in *E. coli* than in *L. casei*. In *L. casei* the Arg-55 site is filled by glycine. Nonetheless, Glu-3 contributes roughly the same binding energy in *L. casei* as in *E. coli*. This suggests that other positively charged residues in the vicinity might contribute to the binding of Glu-3 in *L. casei*. Indeed, electrostatic calculations reveal a region of strong positivity on the surface of *L. casei* TS in exactly the same location as the region in the *E. coli* structure arising from Arg-55 (data

	48	78	304	313
E. coli	--TT**K**RC**H**L**R**SI**I**HELLWFLQGDTNIAYLHEN**N**--		--YDPHPGI**K**AP--	
L. casei	--TT**K**KB**P**F**G**LI**K**SELLWFLHGDTNIRFLLQH**R**--		--YDPYPAI**K**AP--	
Human	--TT**K**RV**F**W**K**GV**L**EELLWFIKGSTNAKELSSK**G**--		--YNPHPTI**K**ME--	
Mouse	--TT**K**RV**F**W**K**GV**L**EELLWFIKGSTNAKELSSK**G**--		--YNPHPTI**K**ME--	
	50 53 55 58			311

FIGURE 6: Alignment of two regions of TS sequences that participate in binding the polyglutamyl moiety. The numbering scheme corresponds to the *L. casei* sequence. The segment including residues 48–78 contains helices B and C; the segment including residues 304–311 is very close to the C-terminus. Residue positions mentioned in the text as contact points for the polyglutamyl moiety are printed in boldface and numbered below.

not shown). This positive field derives from the guanidinium group of *L. casei* Arg-78 that reaches across from the C helix into a position where it can contact Glu-3 (Figure 4). Three residues in linear sequence beyond the Arg-55 site is a lysine (*L. casei* Lys-58) that would be in a position to contact a fifth or sixth glutamate residue if the polyglutamyl moiety continued in the same direction across the protein surface. Lys-58 has been implicated in binding to the polyglutamate moiety by biochemical cross-linking experiments (Maley et al., 1982). This may explain why in *L. casei* Glu-5 and Glu-6 contribute a considerable sum to the overall binding energy, whereas in the other species that lack a basic residue in this locale such as *E. coli*, Glu-5 and Glu-6 do not contribute much binding energy. The mammalian sequences not only lack a basic residue but also possess an additional glutamate residue just one residue in linear sequence beyond. This acidic residue may generate a local negative field that destabilizes the polyglutamyl moiety in this area, thus explaining the reduction in binding energy associated with Glu-5,6 in pig and human TS.

Interaction between TS and the Polyglutamyl Group Is Primarily Electrostatic. The binding mode of the polyglutamyl moiety is different from the typical ligand/receptor interaction in that it involves few highly localized interactions such as hydrogen bonds and van der Waals contacts. Instead, the interaction may depend on more diffuse electrostatic forces that permit greater flexibility and freedom of motion. There are only three hydrogen bonds formed by the moiety that were clearly discernible. All of these hydrogen bonds anchor Glu-1 via water molecules to the protein. The high *B*-factors for the atoms in Glu-2 and Glu-3 and their neighboring protein atoms suggest that the structure is mobile and does not require specific interactions for integrity. The mobility and flexibility of polyglutamate suggest that the binding energy between TS and the polyglutamyl group may be relatively insensitive to the structural details of the TS binding surface. Substitutions in the regions that contact Glu-2 and Glu-3, especially conservative substitutions, should be commonplace. Comparison of several TS sequences shows that only Lys-50 is nearly invariant (17/18) among all TS enzymes. The other residues vary. For example, the Arg-55 position is filled in every case (except *L. casei*) by arginine (9/18) or lysine. The His-53 position is most frequently (11/17) filled by phenylalanine, but it can be substituted with histidine, proline, serine, or alanine.

The binding of the polyglutamyl moiety to TS likely involves a substantial reduction of entropy. Above pH 5.5, polyglutamate peptides exist in aqueous solution as random coils (Rialdi & Hermans, 1966). Therefore, the number of accessible states and hence the entropy of the unbound polyglutamyl moiety are probably high. Because the *B*-factors of surface side-chain atoms that interact with the polyglutamyl group are as high in unbound structures as in the TS/dUMP/CB3717polyglu structure, the overall entropy change in the protein is probably small. Thus the dominant factor in the entropy change associated with cofactor binding is the loss of entropy in the polyglutamyl group itself. An electrostatic binding interaction that demands few specific, highly directional or distance-dependent interactions may be favored in a situation where the entropic costs associated with binding should be high.

The Polyglutamyl Binding Site: An Alternative Inhibitor Target. It is attractive to consider the possibility of using the polyglutamate binding site as a target for inhibitors. If the interaction between the polyglutamyl group and TS could be mimicked and improved, it might be possible to develop a family of TS inhibitors that were more species selective than current inhibitors such as 5-FU (Maley et al., 1986). This follows from the observation that the most highly conserved region of TS is the active site; residues farther from the active site are less conserved (Perry et al., 1990). There is considerable variability in the region that contacts the second and third glutamate residues in the polyglutamyl moiety (Figure 6). Thus it might be possible to capitalize on these differences to construct an inhibitor that recognizes, for example, fungal or protozoal TS but does not bind human TS.

Polyglutamate attached to a benzoyl group is a poor inhibitor of TS (IC_{50} > 2 mM), but the crystal structure of the polyglutamyl moiety bound to TS should assist in design of a more rigid, lower entropy molecule that has a correspondingly lower K_d. High-resolution structural information may be essential to identify regions on the surface of a protein suitable as targets for inhibitors, namely, concave regions with low *B*-factors. In the vicinity of the *E. coli* TS active site, there is no region on the protein surface that has significantly lower *B*-factors than the area that contacts the polyglutamyl moiety (data not shown). However, Glu-2 does not take full advantage of its position within a shallow canyon. It makes contact primarily with residues that form the bottom of the canyon. Several hydrogen bond donors and acceptors on the side of the canyon, such as the hydroxyl groups and amide nitrogen of Thr-80 and Ser-56 and the carbonyl of Gly-309, do not participate in the binding interaction. A larger chemical group in the same location might be better able to contact the canyon walls, thereby contributing significant binding energy.

CONCLUSION

The polyglutamyl group serves several purposes in folate metabolism such as increasing the solubility of folate inside the cell. To compensate for increased hydrophilicity of the cofactor, TS binds the polyglutamate moiety with a positively charged, hydrophilic surface. This electrostatic interaction explains how different species of TS bind polyglutamate with similar affinities despite possessing considerable sequence divergence in the polyglutamyl binding site. This sequence variability could be exploited in the development of more selective TS inhibitors. The TS/dUMP/CB3717polyglu

structure reveals regions near the polyglutamate binding site that do not directly contact the glutamate residues. These regions contain hydrogen bond donors and acceptors in position to interact with novel inhibitors customized to better fill the binding site. An inhibitor molecule can be envisioned that derives some binding energy from conserved sites, such as the PABA ring binding pocket, and additional energy as well as specificity from the region outside the active site.

ACKNOWLEDGMENT

We are grateful to Dr. V. Ramalingam and E. Fauman for help in the initial phase of this work. We thank Drs. F. and G. Maley for generous gifts of *E. coli* TS, ICI for supplying CB3717polyglu, R. Reed for providing a complete amino acid sequence alignment of published TS sequences, J. Newdoll for help with computer graphics, and C. Carreras and Dr. D. Santi for assistance with TS enzyme assays. Figures 4 and 5 were made using Midas Plus written by Conrad Huang, Eric Pettersen, and Greg Couch at the UCSF Computer Graphics Laboratory.

REFERENCES

Appelt, K., Bacquet, R. J., Bartlett, C. A., Booth, C. L. J., Freer, S. T., et al. (1991) *J. Med. Chem. 34*, 1925–1934.

Belfort, M., Maley, G., & Maley, F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1858–1861.

Blum, M., Metcalf, P., Harrison, S. C., & Wiley, D. C. (1987) *J. Appl. Crystallogr. 20*, 235–242.

Brunger, A. T. (1990) X-PLOR, version 2.1, Yale University, New Haven, CT.

Bystroff, C., & Kraut, J. (1991) *Biochemistry 30*, 2227–2239.

Connolly, M. L. (1983) Analytical Molecular Surface Calculation, *J. Appl. Crystallogr. 16*, 548–558.

Davies, J. F., Delcamp, T. J., Prendergast, N. J., Ashford, V. A., Freisheim, J. H., & Kraut, J. (1990) *Biochemistry 29*, 9467–9479.

Dolnick, B. J., & Cheng, Y.-C. (1978) *J. Biol. Chem. 253*, 3563–3567.

Finer-Moore, J. S., Montfort, W. R., & Stroud, R. M. (1990) *Biochemistry 29*, 6977–6986.

Friedkin, M., Plante, L., Crawford, E. J., & Crumm, M. (1975) *J. Biol. Chem. 250*, 5614–5621.

Hardy, L. W., Finer-Moore, J. S., Montfort, W. R., Jones, M. O., Santi, D. V., & Stroud, R. M. (1987) *Science 235*, 448–455.

Harrap, K. R., Jackman, A. L., Newell, D. R., Taylor, G. A., Hughes, L. R., & Calvert, A. H. (1989) *Adv. Enzyme Regul. 29*, 161–179.

Henderson, G. B. (1990) *Annu. Rev. Nutr. 10*, 319–335.

Hoffbrand, A. V., Tripp, E., Houlihan, C. M., & Scott, J. M. (1973) *Blood 42*, 141–146.

Honig, B., Sharp, K. A., & Gilson, M. (1989) *Prog. Clin. Biol. Res. 289*, 65–74.

Howard, A. J., Nielson, C., & Xuong, Ng H. (1985) *Methods Enzymol. 114*, 452–472.

Lewis, G. P., Salem, M. E., & Rowe, P. B. (1979) in *Chemistry and Biology of Pteridines* (Kisliuk, R. L., & Brown, G. M., Eds.) pp 441–446, Elsevier-North Holland, New York.

Lu, Y.-Z., Aiello, P., & Matthews, R. (1984) *Biochemistry 23*, 6870–6876.

Maley, G. F., Maley, F., & Baugh, C. M. (1982) *Arch. Biochem. Biophys. 216*, 551–558.

Maley, F., Chu, F. K., West, D. K., & Maley, G. F. (1986) in *Chemistry and Biology of Pteridines* (Cooper, B. A., & Whitehead, V. M., Eds.) pp 613–629, Elsevier-North Holland, New York.

Matthews, D. A., Villafranca, J. E., Janson, C. A., Smith, W. W., Welsh, K., & Freer, S. (1990) *J. Mol. Biol. 214*, 937–948.

Matthews, R. G., Chandralekha, G., Green, J. M., Matthews, K. D., & Dunlap, R. B. (1987) *Adv. Enzyme Regul. 26*, 157–171.

Montfort, W. R., Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Maley, G. F., Hardy, L., Maley, F., & Stroud, R. M. (1990) *Biochemistry 29*, 6964–6977.

Nahas, A., Nixon, P. F., & Bertino, J. R. (1972) *Cancer Res. 32*, 1416–1421.

Oefner, C., D'Arcy, A., & Winkler, F. K. (1988) *Eur. J. Biochem. 174*, 377–385.

Paquin, J., Baugh, C. M., & MacKenzie, R. E. (1985) *J. Biol. Chem. 260*, 14925–14931.

Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Montfort, W. R., Maley, G. F., Maley, F., & Stroud, R. M. (1990) *Proteins 8*, 315–333.

Pogolotti, A. L., Jr., Danenberg, P. V., & Santi, D. V. (1986) *J. Med. Chem. 29*, 478–482.

Rialdi, G., & Hermans, J., Jr. (1966) *J. Am. Chem. Soc. 88*, 5719–5720.

Schirch, V., & Strong, W. B. (1989) *Arch. Biochem. Biophys. 269*, 371–380.

Sikora, E., Jackman, A. L., Newell, D. R., & Calvert, A. H. (1988) *Biochem. Pharmacol. 37*, 4047–4054.

Thorndike, J., Kisliuk, R. L., Gaumont, Y., Piper, J. R., & Nair, M. G. (1990) *Arch. Biochem. Biophys. 277*, 334–341.