Conformational Analysis of Thymidylate Synthase from Amino Acid Sequence and Circular Dichroism†

Parthasarathy Manavalan,‡ Denice M. Mittelstaedt, Michael I. Schimerlik, and W. Curtis Johnson, Jr.*
Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331
Received May 5, 1986; Revised Manuscript Received June 16, 1986

ABSTRACT: Circular dichroism studies were carried out in the vacuum ultraviolet region for thymidylate synthase from *Lactobacillus casei* and its ligand complexes. The CD spectrum was analyzed for secondary structure by our method and the variable selection method, and both gave similar results. Our method predicts 33% α-helix, 25% (23% antiparallel and 2% parallel) β-sheet, 20% turns, and 16% other structure. The secondary structure of this protein was also predicted from the amino acid sequence by four different methods. Though there is a variation in the prediction among these methods, the prediction of 32% α-helix and 23% β-sheet by combining the four methods is in excellent agreement with our CD results. Further, the location of the predicted regions of α-helices and β-strands along the sequence and the CD characteristics strongly suggest that this protein belongs to an $\alpha+\beta$ structural class. Binding of the inhibitor FdUMP or the cofactor 5,10-methylenetetrahydrofolate did not change the CD spectrum. However, when both ligands were present, there was a significant change in the CD spectrum and the maximum changes occurred when the concentration of FdUMP was 1 mol/mol of enzyme. The addition of FdUMP and cofactor causes, respectively, a 5% and 6% decrease in β-sheet and β-turns and about an 8% increase in "other" structure.

Thymidylate synthase catalyzes the reaction involving the conversion of deoxyuridylate (dUMP) to thymidylate (dTMP). This reaction is a two-step process in which transfer of a one-carbon unit from the cofactor N^5,N^{10}-methylenetetrahydrofolate (5,10-CH_2H_4folate) to the 5-position of dUMP is followed by reduction of the one-carbon unit to a methyl group (Lomax & Greenberg, 1967). Because of the role that thymidylate synthase plays in DNA synthesis, it has been an important chemotherapeutic target enzyme in the treatment of various proliferative diseases (Reyes & Heidelberger, 1965).

Thymidylate synthase has been isolated from different sources and the complete amino acid sequence for *Lactobacillus casei* (*L. casei*) has been determined by various cleavage techniques and peptide analysis (Maley et al., 1979). Sequences for this protein from *Escherichia coli* and bacteriophage T4 have also been predicted by the gene-cloning technique (Belfort et al., 1983; Chu et al., 1984). The enzyme from *L. casei* has a monomer molecular weight of 35 000 and is composed of two identical subunits (Dunlap et al., 1971).

It has been found that FdUMP, a fluorinated analogue of dUMP, is a potent inhibitor of thymidylate synthase (Dannenberg et al., 1958). The enzyme has two identical sites, and in the presence of cofactor 5,10-CH_2H_4folate, FdUMP binds extremely tightly (K_d = 10^{-11}–10^{-13} M) to both (Lewis & Dunlap, 1981). Further, kinetic binding studies with ligands of different concentrations suggest a ligand-induced sequential model for subunit interactions, in which only one site is accessible on the free enzyme until the binding of both ligands, 5,10-CH_2H_4folate and FdUMP, causes a conformational change that opens the second internal site (Danenberg & Danenberg, 1979). The determination of the three-dimensional structure of this enzyme will greatly help in understanding the mechanism involved in ligand binding. In the absence of crystal structure data, empirical prediction methods to identify the secondary structural regions from the amino acid sequence and solution studies of secondary structure by circular dichroism (CD) can be used to elucidate the structure of the proteins (Chou & Fasman, 1978; Johnson, 1984).

In this paper, we have undertaken CD studies of thymidylate synthase from *L. casei* to probe the structural details of the enzyme and conformational changes caused by ligand binding. Since the sequence is also available for this enzyme (Maley et al., 1979), we also predict the secondary structure from the amino acid sequence. We use four different prediction methods (Burgess et al., 1974; Chou & Fasman, 1978; Lim, 1974; Garnier et al., 1978) and combine the results to produce a joint prediction of the α-helices and β-sheet regions. CD measurements are made from 260 to 178 nm for thymidylate synthase and its ligand complexes and the results analyzed for structural information. For the secondary structure analysis of CD data, we use the method developed in our laboratory and a new "variable selection" method.

Analysis of secondary structure for thymidylate synthase indicates that the average amount of α-helix and β-sheet predicted from the amino acid sequence is in good agreement with the CD estimation. Further, the clustering of α-helices and β-strands, which are segregated along the sequence, suggests that the tertiary folding pattern is of the $\alpha+\beta$ type. The CD characteristics of this protein also predict the $\alpha+\beta$ classification. No significant changes in the CD spectrum are found for the binding of dUMP, FdUMP, or cofactor, but a definite change occurs when both FdUMP and cofactor are bound to the enzyme.

Materials and Methods

Materials

Thymidylate synthase (sp act. 3.4 μmol min^{-1} mg^{-1}) was purified from an amethopterin-resistant strain of *L. casei* according to the procedure of Dunlap et al. (1971) as modified

† This work was supported in part by NIH Grant GM-21479 (W.C.J.) and grants from the Oregon Medical Research Foundation and the Oregon Chapter of the American Cancer Society (M.I.S.). M.I.S. was the recipient of Research Career Development Award HL 00796 from NIH.
‡ Present address: Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60680.

0006-2960/86/0425-6650$01.50/0

by Galivan et al. (1975) in the presence of 20 mM β-mercaptoethanol. Purity of the enzyme was determined by 7.5% SDS–polyacrylamide gels, which yielded one band at 35 000 Da. The enzyme was stored at 0 °C in saturated ammonium sulfate, 0.05 M potassium phosphate, and 20 mM β-mercaptoethanol, pH 7.1. Prior to use, the enzyme was dialyzed with two changes of 1 L of 0.01 M potassium phosphate and 5 mM β-mercaptoethanol, pH 7.1, under argon (for protection of the active cysteine) at 4 °C.

The nucleotide dUMP was purchased from Sigma, while FdUMP and the (+)-*l*-methylenetetrahydrofolate (5,10-CH_2H_4folate) isomer were the generous gifts of Dr. Peter Danenberg. All experiments were performed in 0.01 M potassium phosphate and 5 mM β-mercaptoethanol, pH 7.1. Ligands were added from concentrated aliquots in the same buffer and were used at concentrations between 25 and 50 times that of the protein. Also, changes in the CD were monitored at 208 nm for different molar ratios of FdUMP and thymidylate synthase in the presence of a 5-fold molar excess of 5,10-CH_2H_4folate.

Methods

Secondary Structure Prediction from Amino Acid Sequence. The amino acid sequence of thymidylate synthase from *L. casei* was taken from Maley et al. (1979). The prediction methods of Chou and Fasman (1978), Burgess et al. (1974), Lim (1974), and Garnier et al. (1978) were used to estimate α-helices and β-strands. A computer program was developed for the Chou and Fasman method, and the average helical and β-sheet potentials of tetrapeptide i to $(i + 3)$ were calculated from the respective P_α and P_β values of single residues (Chou & Fasman, 1978). Prediction rules of Burgess et al. (1974) have been slightly modified; a residue is in α (or β) whenever its prediction parameter is greater than that for β (or α) and turn. This maximizes α-helical and β-sheet regions. For the Robson method, we used decision constants −100.0 and −87.5 for α-helix and extended structures, respectively, which are recommended when the secondary structure is more than 50% (Garnier et al., 1978). The following rules were applied for the joint prediction of secondary structures from these four different methods:

(1) The secondary structure was predicted by using individual methods.

(2) A residue was assigned α-helix or β-sheet conformation if it was predicted by a majority of the methods. To be considered α-helix, a minimum of four consecutive residues must be in the α-helical conformation; to be considered β-strand, a minimum of three consecutive residues must be in the β-sheet conformation.

(3) If any overlap of secondary structure prediction occurs (say two predicting α-helix and two predicting β-sheet), the method of Kabat and Wu (1973) and hydrophobicity/hydrophilicity criteria (Kanehisa & Tsong, 1980; Cid et al., 1982) were used to clarify the ambiguity.

CD Measurements. CD measurements were made on a vacuum UV CD spectrophotometer at 20 ± 1 °C with 50–100-μm path length cells. The instrument was calibrated with (+)-10-camphorsulfonic acid, $\Delta\epsilon = +2.37$ M^{-1} cm^{-1} at 290.5 nm. Transmission spectra were also measured with the same instrument. The extinction coefficient determined at 190 nm is 9500 ± 200 M^{-1} cm^{-1}. The entire procedure was repeated 6 times for the sample prepared at different times with different concentrations. The concentration of the sample varied from 0.5 to 1.5 mg/mL. Data were collected by a microcomputer with a 0.1-s time constant at a scanning rate of four data points per minute with a 0.5-nm interval. Before the base line was subtracted, the digitized data for the sample and the buffer were smoothed with 11- and 19-point Savitzky–Golay smoothing functions, respectively (Savitzkey & Golay, 1964). Concentration of the protein was determined from the ninhydrin reaction common to amino acid analysis but without separating the residues, as described by Hennessey and Johnson (1981).

Analysis of CD Data for Secondary Structure. The digitized CD data were analyzed for secondary structure by different methods. Forty-two data points at 2-nm intervals between 178 and 260 nm were considered for all analyses. Besides using the method developed in our laboratory (Hennessey & Johnson, 1981), we also used the "variable selection" method (Mosteller et al., 1977), both of which are briefly outlined here.

Method of Hennessey and Johnson. Hennessey and Johnson (1981) applied singular value decomposition (SVD) to the CD data for 16 reference proteins of known secondary structure. They used the five most important basis vectors from SVD to predict the secondary structure of a protein from its CD spectrum. This method was further improved as a more simple and straightforward procedure by Compton and Johnson (1986). According to this method the linear relationship between CD data and their corresponding secondary structure can be written as the matrix equation

$$\mathbf{XC} = \mathbf{F} \tag{1}$$

Here, **C** are the CD data of the 16 reference proteins, **F** is the corresponding fraction of the protein secondary structure, and **X** are the coefficients that relate C and F. **X** can be determined by transforming eq 1 as

$$\mathbf{X} = \mathbf{FC}^{-1} \tag{2}$$

where $\mathbf{C}^{-1}$ is the inverse of **C**. The matrix **C** can be decomposed by SVD into the product $\mathbf{USV}^T$ where **U** and **V** are orthogonal and unitary matrices and **S** is a matrix with entries only of the main diagonal. The elements of **S** are called singular values. Then

$$\mathbf{C}^{-1} = \mathbf{VS}^+\mathbf{U}^T \tag{3}$$

where $\mathbf{S}^+$ has nonzero entries only on the main diagonal that are the reciprocals of the singular values (Forsythe et al., 1977). Multiplying a digitized CD spectrum by **X** will give the secondary structure of the corresponding protein. In calculating $\mathbf{C}^{-1}$, we use only the five most significant singular values and the corresponding vectors in **U** and **V** matrices to eliminate noise and avoid instability.

Variable Selection Method. The variable selection method consists of removing one or more references that contain extraneous variables from the data used to perform the analysis (Mosteller & Tukey, 1977). In our case, this is equivalent to removing proteins that have extraneous contributions to the CD from the basis set and recalculating the secondary structure. By removing CD contributions due to factors that are not found in the protein being analyzed, we hope to have no more variables than can be handled by the information content of our CD data. Since we do not know a priori what proteins should be removed from the basis set, calculations were performed for all possible subsets. If there are n proteins in the basis set and r proteins to be removed, there are $n!/r!(n - r)!$ combinations. The criteria for selecting the best possible solution(s) among the subsets depend upon the nature of the problem. For our analyses, the following criteria have been applied to choose the optimum solutions: (i) The sum of the predicted secondary structure should be in the range 0.9–1.10. (ii) No secondary structure should be large and negative. (iii) The fit of the reconstructed CD should be within the noise level

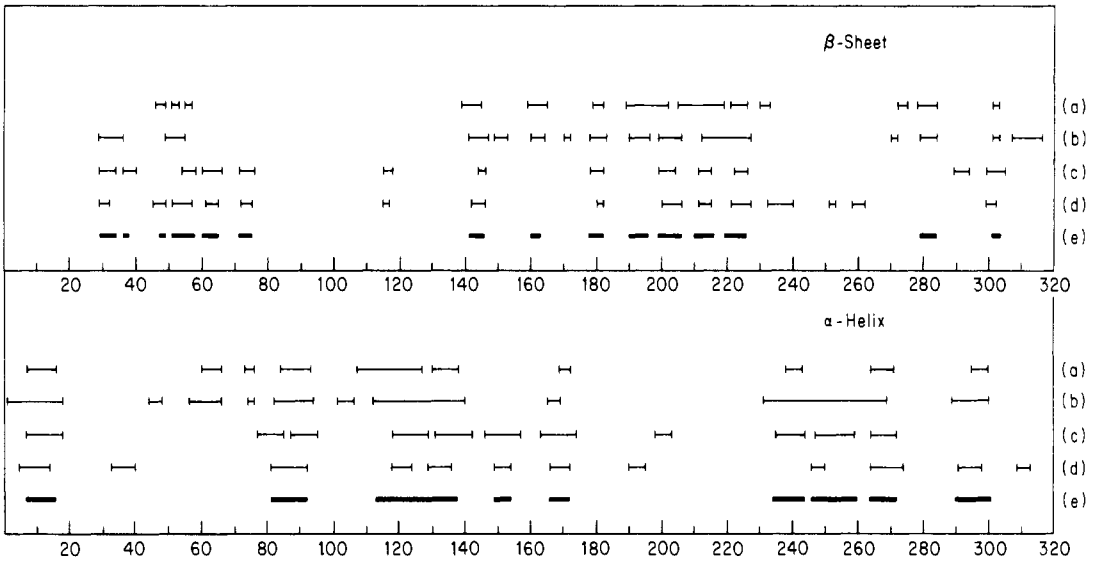

FIGURE 1: Prediction of α-helices and β-strands from amino acid sequence by different methods: (a) Burgess et al. (1974); (b) Garnier et al. (1978); (c) Lim (1974); (d) Chou and Fasman (1978); (e) joint prediction.

Table I: Secondary Structure Analysis of Thymidylate Synthase from Amino Acid Sequence

method	α-helix	β-sheet (A + P)[a]	other (T + O)[b]
Chou and Fasman (per residue)	0.41	0.28	0.31
Chou and Fasman (per amide)	0.37	0.24	0.39
Robson group (per residue)	0.41	0.33	0.26
Robson group (per amide)	0.38	0.28	0.34
Lim (per residue)	0.37	0.22	0.41
Lim (per amide)	0.33	0.18	0.49
Burgess et al. (per residue)	0.26	0.21	0.53
Burgess et al. (per amide)	0.23	0.17	0.60
joint method (per residue)	0.34	0.27	0.39
joint method (per amide)	0.32	0.23	0.45

[a] A, antiparallel; P, parallel. [b] T, β-turns; O, other structures.

of the experimental data. Analyses of the proteins in the basis set indicate that the noise level is equivalent to the sum of the squares of the residuals (SSR) being less than 2.0. (iv) The largest basis set that gives reasonable results should be used. (v) All proteins in the basis set that have a CD similar to the test protein should be included. Since the CD of lysozyme is similar in shape to the CD of thymidylate synthase, we considered only those subsets that included lysozyme.

Results and Discussion

Secondary Structure Analysis from Amino Acid Sequence. Table I gives the amount of each secondary structure predicted from amino acid sequence by the four different methods. From Figure 1 we see that the regions around the segments 7–16, 81–92, 113–138, 166–172, 234–244, 264–272, and 290–300 are predicted as α-helices by three out of four or by all four methods. Similarly, the commonly predicted regions for β-strands are 29–34, 51–58, 141–146, 178–182, 199–207, 210–216, 219–226, and 301–303 (Figure 1). Although all the methods predict most of the α-helices and β-strands in similar regions along the sequence, there is a considerable variation in their length and hence in the total amount in each category. The prediction of α-helical content by all the methods except Burgess et al. (1974) is in good agreement (37–41%). However, there is a large difference in the prediction of β-sheet, varying from 21% to 33%. The Robson method (Garnier et al., 1978) predicts the highest amount, whereas the lowest value is predicted by Burgess et al. (1974). When we changed the decision constants that correspond to total secondary structure in the Robson method for the range 20–50%, the changes in the prediction of secondary structure were insignificant.

Since the joint prediction algorithm seems to give more reliable results than any single prediction method, we applied this procedure (as outlined under Methods) to estimate the secondary structure regions. The locations of α-helices and β-strands predicted by this method are given in Table II, and a schematic of these predicted regions along the sequence is shown in Figure 2.

Table II also includes the β-turns predicted by the Chou and Fasman method (1979). It is obvious from the figure and table that α-helical regions are dominant in the latter part of the chain, whereas β-strands are concentrated in the N terminus (29–75) and middle portion (178–226). It is interesting to note that the segment 113–138, which is not present in the

Table II: Location of α-Helices, β-Strands, and β-Turns in Thymidylate Synthase (*L. casei*)[a]

α-helices	β-strands	β-turns
7–16, 81–92, 113–138, 149–154, 166–172, 234–244, 246–260, 264–272, 290–300	29–34, 47–49, 51–58, 60–65, 71–75, 141–146, 160–163, 178–182, 190–196, 199–207, 210–216, 219–226, 279–284, 301–303	15–18, 20–23, 25–28, 41–44, 66–69, 68–71, 77–80, 93–96, 98–101, 102–105, 106–109, 109–112, 111–114, 137–140, 139–142, 147–150, 154–157, 156–159, 162–165, 174–177, 176–179, 186–189, 196–199, 207–210, 216–219, 242–245, 275–278, 285–288, 287–290, 303–306, 305–308

[a] α-Helices and β-strands have been predicted by the joint prediction algorithm and β-turns by the Chou and Fasman method (Chou & Fasman, 1979).

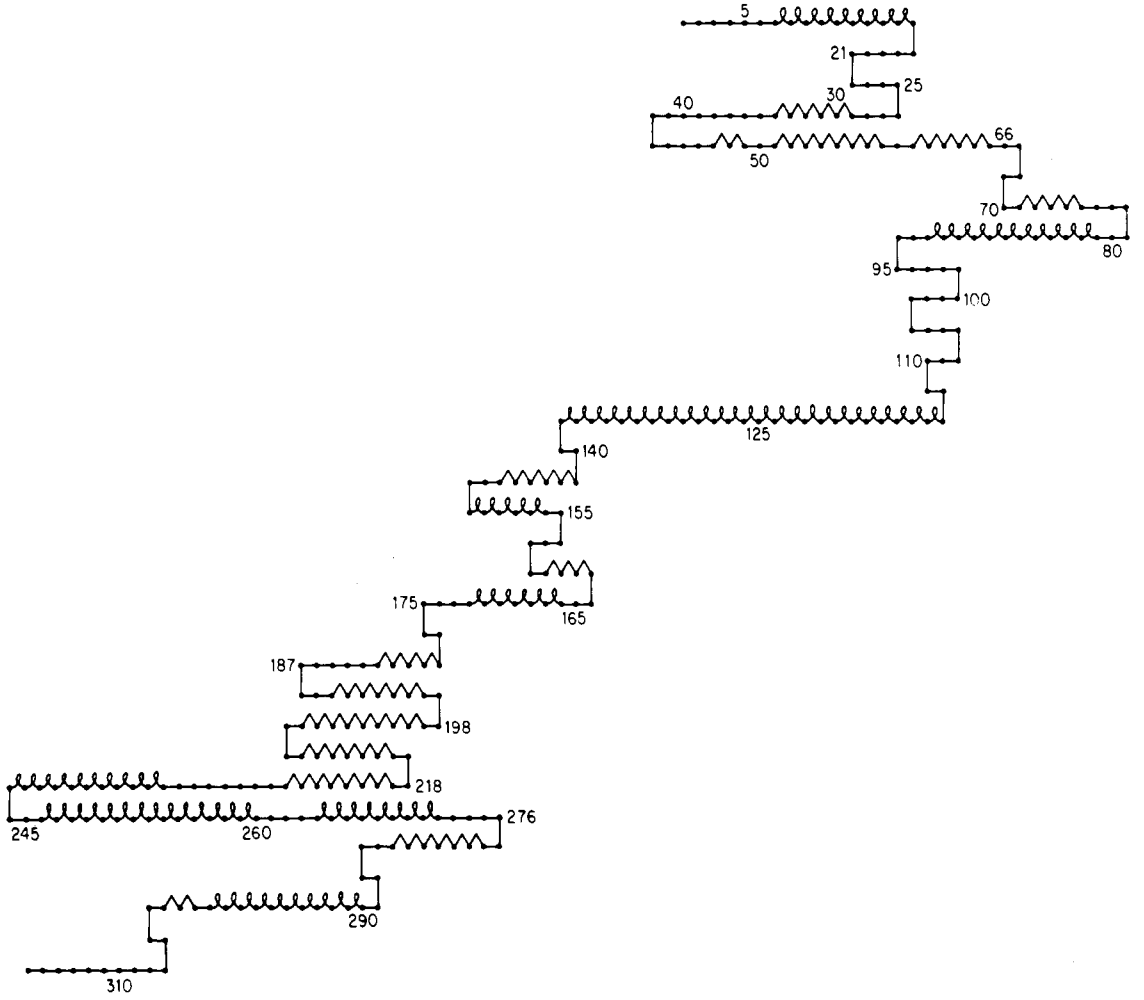

FIGURE 2: Schematic diagram of the predicted secondary structure in thymidylate synthase. α-Helices (coiled) and β-strands (jagged) are from the joint prediction algorithm and β-turns from the Chou and Fasman method (Table II).

Table III: Secondary Structure for Analyzing CD Data of Thymidylate Synthase and Its Ligand Complex[a]

method	H	A	P	T	O	total
CD of thymidylate synthase						
Hennessey and Johnson (unconstrained)	0.33 ± 0.03	0.23 ± 0.02	0.02 ± 0.01	0.20 ± 0.03	0.16 ± 0.05	0.94 ± 0.08
variable selection method (unconstrained)	0.33 ± 0.00	0.24 ± 0.01	0.02 ± 0.00	0.21 ± 0.01	0.20 ± 0.00	1.00 ± 0.00
CD of thymidylate synthase + FdUMP + 5,10-CH_2H_4folate						
Hennessey and Johnson (unconstrained)	0.31 ± 0.02	0.16 ± 0.02	0.02 ± 0.01	0.17 ± 0.03	0.18 ± 0.03	0.84 ± 0.05
variable selection method (unconstrained)	0.32 ± 0.01	0.19 ± 0.00	0.04 ± 0.01	0.15 ● 0.00	0.28 ± 0.03	0.98 ± 0.02

[a] H = α-helix, A = antiparallel β-sheet, P = parallel β-sheet, T = β-turns, and O = others.

bacteriophage thymidylate synthase sequence, is predicted as a long α-helix. Finally, the secondary structure estimated by the joint prediction algorithm amounts to 34% α-helix, 27% β-sheet, and 39% other structure (Table I).

Secondary Structure Analysis from CD Data. The CD spectrum depicted in Figure 3 is an average of six measurements. The error bars at 193, 208, and 220 nm are the deviations from the average value and hence represent the repeatability of $\Delta\epsilon$ for our measurements. The shape of the CD is similar for all runs. The spectrum is characterized by two negative minima around 208 and 220 nm and a positive peak at 193 nm. The spectrum crosses the base line from negative to positive at 200–201 nm and again from positive to negative at 178–179 nm. As described under Methods, we computed the **X** from CD spectra of proteins with known secondary structure. Multiplication of the CD of thymidylate synthase by **X** gives the estimation of each secondary structure. The secondary structure analysis for the CD spectrum shown in Figure 3 yields 33% α-helix, 23% antiparallel and 2% parallel

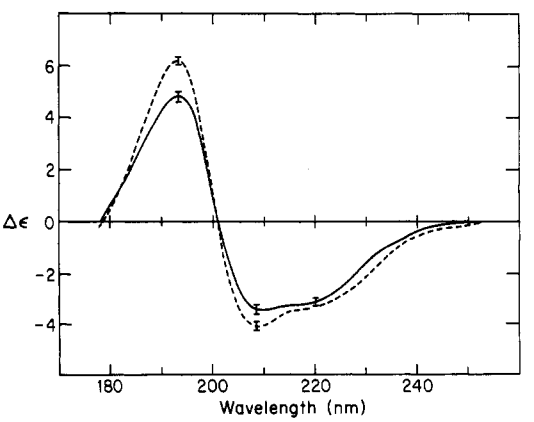

FIGURE 3: CD spectra of thymidylate synthase (---) and as a complex with ligands FdUMP and 5,10-CH_2H_4folate (—).

β-sheet, 20% β-turns, and 16% "other" structure, which totals 94% (Table III). Also, the reconstruction of the CD spectrum

is excellent, having an SSR = 0.520.

Since we do not apply any constraint in our method of analysis, it is possible that the total may not add up to 1.00, and the fractions of secondary structure may have negative values. These discrepancies may be due to the proteins with CD contributions that are not found in the protein being analyzed. Removal of these proteins from the basis set should greatly improve the analysis for secondary structure. This variable selection is not new and is a standard procedure in the statistical analysis of data (Mosteller & Tukey, 1977). When we applied the new procedure, as outlined under Methods, for each of the 16 proteins included in the basis set, our estimations of β-sheet, β-turns, and other structure were very much improved. The correlation coefficients of H, A, P, T, and O with all the 16 proteins in the basis set are 0.99, 0.77, 0.79, 0.84, and 0.82, respectively, whereas with our new procedure they are 0.99, 0.92, 0.89, 0.90, and 0.91. A more thorough analysis on the prediction of secondary structures using the variable selection procedure is in preparation.

In the case of thymidylate synthase, the analysis using all 16 proteins itself is good, as it fits the CD within the noise level and the total adds to 0.94. The variable selection method becomes important when we analyze the CD of thymidylate synthase complexed with ligands. However, to be consistent, we applied it here to uncomplexed thymidylate synthase and obtained three subsets, with the sum of the secondary structure equal to 1.00. These subsets have 14 proteins in the basis set and satisfied all the criteria mentioned in the Methods section. The average values are given in Table III and indicate that all the secondary structures have similar values to those obtained with the original method. Only the other structure has been increased from 16% to 20%, which again reflects that in most cases the error tends to accumulate in the other structure (Hennessey & Johnson, 1982).

The amide, not the residue, in each ordered segment contributes to the CD. Thus, for purposes of comparison with CD results, the fractions of secondary structure predicted from sequence are based on the number of amides instead of residues. For each segment of predicted α-helix or β-sheet, the number of amides is one fewer than the number of residues. Also, the entire protein has one fewer amide than the total number of residues. Basing the sequence prediction on amides can make a sizable change [e.g., for a protein having l segments of α-helices, m residues in these segments, and n total residues in the protein, the fraction of α-helix in terms of amides is $(m - l)/(n - 1)$]. If we change the sequence prediction values accordingly, the joint prediction results in terms of amides should be 32% α-helix and 23% β-sheet. These results agree very well with our CD estimation. Also, the predictions of α-helix and β-sheet by Chou and Fasman and by the Robson group are very close to CD estimation, while α-helix prediction by Lim agrees with the CD results (Tables I and III).

CD Analysis of Thymidylate Synthase and Ligand Complexes. We also used CD to monitor the conformational changes of thymidylate synthase upon binding of ligands. The addition of dUMP, FdUMP, or cofactor 5,10-CH_2H_4folate did not change the CD spectrum measured. However, when FdUMP was added to thymidylate synthase in the presence of cofactor, we noticed a definite change in the CD spectrum of thymidylate synthase (Figure 3). The addition of both ligands decreased the magnitude of the positive band at 193 nm as well as the negative band at 208 nm. The titration of FdUMP with thymidylate synthase in the presence of a 5-fold molar excess of cofactor revealed that the maximum changes in magnitude occur when the concentration of FdUMP is 1 mol/mol of enzyme. The conformational changes, studied by hydrodynamic methods (Lockshin & Danenberg, 1980), also reveal that 70% of the change in Stokes' radius occurs upon binding of 1 mol of ligands/mol of enzyme. Furthermore, Sharma and Kisliuk (1975) have studied the titration of FdUMP by fluorescence techniques and report that maximum quenching occurs when 1 mol of FdUMP interacts with 1 mol of enzyme. These results strongly suggest that there is some kind of structural alteration when the first site is occupied by FdUMP and cofactor.

The CD spectrum of the complex shown in Figure 3 was also analyzed for secondary structure. As Table III shows, the analysis by the Hennessey and Johnson method gives a total equal of 0.84. However, with the variable selection method we obtained a total value of 0.98. Again, the increase in total from 0.84 to 0.98 occurs mostly in the other structure increasing its value from 0.18 to 0.28. When these results are compared with the analysis of thymidylate synthase obtained by the variable section method, we find a definite structural change in the thymidylate synthase–ligand complex. Thus, according to this analysis, addition of FdUMP and cofactor causes, respectively, about a 5% and 6% decrease in β-sheet and β-turns and about an 8% increase in other structure. It is of interest to note here that cysteine residue 198 is involved in FdUMP binding and is part of a β-turn that connects two β-strands (Figure 2).

Structural Domains in Thymidylate Synthase. It is obvious from the analyses for secondary structure that thymidylate synthase is a mixture of α-helices and β-strands. Also, α-helices and β-strands along the chain do not occur alternatively as observed in α/β type proteins (Levitt & Chothia, 1976), but each structure seems to cluster in separate regions of the molecule (Figure 2). This segregation of α-helices and β-strands suggests the $\alpha+\beta$ tertiary structural type. Also, the larger magnitude of the 220-nm CD band, a positive peak at 193 nm, and crossover at 178–179 nm are all characteristics of $\alpha+\beta$ structural type (Manavalan & Johnson, 1983), thus agreeing with the sequence prediction. Further, the monomer molecular weight of 35 000 Da puts this protein in the two-domain classification (Manavalan & Johnson, 1985), in which case one domain may be predominantly α-helical and the second domain predominantly antiparallel β-sheet.

ACKNOWLEDGMENTS

We thank Dr. J. Garnier for sending his computer program.

Registry No. FdUMP, 134-46-3; 5,10-CH_2H_4folate, 3432-99-3; thymidylate synthase, 9031-61-2.

REFERENCES

Belfort, M., Maley, G., Lane, J. P., & Maley, F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4914–4918.

Burgess, A. W., Ponnuswamy, P. K., & Scheraga, H. A. (1974) *Isr. J. Chem. 12*, 239–286.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Chou, P. Y., & Fasman, G. D. (1979) *Biophys. J. 26*, 367–384.

Chu, F. K., Maley, G. F., Maley, F., & Belfort, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3049–3053.

Cid, H., Bunster, M., Arriagada, E., & Campos, M. (1982) *FEBS Lett. 150*, 247–254.

Compton, L., & Johnson, W. C., Jr. (1986) *Anal. Biochem. 155*, 155–167.

Danenberg, K. D., & Danenberg, P. V. (1979) *J. Biol. Chem. 254*, 4345–4348.

Dannenberg, P. B., Montag, B. J., & Heidelberger, C. (1958) *Cancer Res. 18*, 329–334.
Donato, H., Jr., Aull, J. L., Lyon, J. A., Reinsch, J. W., & Dunlap, R. B. (1976) *J. Biol. Chem. 251*, 1303–1310.
Dunlap, R. B., Harding, N. G. L., & Huennekens, F. M. (1971) *Biochemistry 10*, 88–97.
Forsythe, G. E., Malcolm, M. A., & Moler, G. B. (1977) in *Computer Methods for Mathematical Computations*, Prentice-Hall, Englewood Cliffs, NJ.
Galivan, J. H., Maley, G. F., & Maley, F. (1975) *Biochemistry 14*, 3338–3344.
Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.
Hennessey, J. P., Jr., & Johnson, W. C., Jr. (1981) *Biochemistry 20*, 1085–1094.
Hennessey, J. P., Jr., & Johnson, W. C., Jr. (1982) *Anal. Biochem. 125*, 177–188.
Johnson, W. C., Jr. (1984) in *Food Analysis: Principles and Techniques* (Gruenwedel, D. W., & Whitaker, J. R., Eds.) Vol. 2, pp 245–291, Marcel Dekker, New York.
Kabat, E. A., & Wu, T. T. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 1473–1477.
Kanehisa, M. I., & Tsong, T. Y. (1980) *Biopolymers 19*, 1617–1628.
Levitt, M., & Chothia, C. (1976) *Nature (London) 261*, 552–557.
Lewis, C. A., Jr., & Dunlap, R. B. (1981) in *Topics in Molecular Pharmacology* (Burgen, A. S. V., & Roberts, G. C. K., Eds.) Vol. 1, pp 169–219, North-Holland Biomedical, New York.
Lewis, C. A., Jr., Ellis, P. D., & Dunlap, R. B. (1981) *Biochemistry 20*, 2275–2285.
Lim, V. I. (1974) *J. Mol. Biol. 88*, 873–894.
Lockshin, A., & Danenberg, P. V. (1980) *Biochemistry 19*, 4244–4251.
Lomax, M. I. S., & Greenberg, R. G. (1967) *J. Biol. Chem. 242*, 1302–1306.
Maley, G. F., Bellisario, R. L., Guarino, D. W., & Maley, F. (1979) *J. Biol. Chem. 254*, 1301–1304.
Manavalan, P., & Johnson, W. C., Jr. (1983) *Nature (London) 305*, 831–832.
Manavalan, P., & Johnson, W. C., Jr. (1985) *Biochim. Biophys. Acta 829*, 365–370.
Mosteller, F., & Tukey, J. W. (1977) in *Data Analysis and Regression*, Addison-Wesley, Reading, MA.
Reyes, P., & Heidelberger, C. (1965) *Mol. Pharmacol. 1*, 14–30.
Savitzky, A., & Golay, A. E. (1964) *Anal. Chem. 36*, 1627–1639.
Sharma, R. K., & Kisliuk, R. L. (1975) *Biochem. Biophys. Res. Commun. 64*, 648–655.

Characterization of the Phosphoserine of Pepsinogen Using ^{31}P Nuclear Magnetic Resonance: Corroboration of X-ray Crystallographic Results†

Susan P. Williams,‡ William A. Bridger, and Michael N. G. James*
Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7
Received March 20, 1986; Revised Manuscript Received May 29, 1986

ABSTRACT: The endogenous phosphoserine residue in porcine pepsinogen has been titrated with use of phosphorus-31 nuclear magnetic resonance (^{31}P NMR). It has an observed pK_{a_2} of 6.7 and a narrow line width ($\simeq$10 Hz). The phosphate can be readily removed by an acid phosphatase from potato; however, it is resistant to hydrolysis by several alkaline phosphatases. The X-ray crystal structure of porcine pepsinogen at 1.8-Å resolution [James, M. N. G., & Sielecki, A. (1986) *Nature (London) 319*, 33–38] shows a rather weak and diffuse region of electron density in the vicinity of the phosphorylated serine residue. This suggests considerable dynamic mobility or conformational disorder of the phosphate. In order to define more fully this behavior, the NMR data have been used to corroborate these crystallographic results. All these physical data are consistent with a highly mobile phosphoserine residue on the surface of the zymogen and freely exposed to solvent. In addition, certain properties of this phosphoserine moiety on pepsinogen are similar to those of one of the phosphorylated residues of ovalbumin. The possible significance of this is discussed.

Both pepsin and pepsinogen have long been known to contain one atom of phosphorus per molecule (Northrop, 1939), this being present as phosphoserine (Flavin, 1954). Enzymatic removal of this phosphate group has no effect on the activity of either pepsin or pepsinogen (Perlmann, 1952), and its function is still unknown. Early studies on the enzymic dephosphorylation of pepsinogen (Perlmann, 1955, 1958) suggested that the phosphate was present in a diester linkage (i.e., linking together two regions of the protein, as, for example, does a disulfide bridge). Later studies on the kinetics of dephosphorylation using a potato phosphatase (Clement et al., 1970) led these workers to favor a phosphate monoester moiety. This was further confirmed by ^{31}P NMR (Edmondson & James, 1979). However, the only evidence provided for this were the chemical shifts of the phosphoryl group of pepsinogen at two pH extremes. These values were quoted to be identical with those of free *O*-phosphoserine.

† This work was supported by a grant to the Medical Research Council of Canada Group in Protein Structure and Function and by the Alberta Heritage Foundation for Medical Research.
‡ Postdoctoral Fellow of the Alberta Heritage Foundation for Medical Research.

0006-2960/86/0425-6655$01.50/0