produced by the radioisotope depends primarily on the distance to the target atom and is not affected by protein-induced alterations in chemical reactivity of this atom. Importantly, the nucleic acids and proteins are as 'transparent' for the low energy Auger electrons as water. Therefore, radioprobing, like NMR, allows one to obtain information on interatomic distances and, in principle, to reconstruct the three-dimensional structure of nucleic acids in complexes with proteins. Radioprobing offers considerable promise as a new method for structural biology in that it requires only a small amount of material, does not demand special equipment and can be applied to large nucleoprotein complexes.

Methods

DNA substrates. Oligonucleotides were synthesized and purified as described[9]. The primer oligonucleotide was labeled with [γ-^{32}P]-ATP using T4 polynucleotide kinase (New England Biolabs), and the template oligonucleotide was labeled with [α-^{32}P]-dATP using terminal transferase from calf thymus (Boehringer Mannheim) following the manufacturer's protocols. The ^{32}P-labeled oligonucleotides were purified on a MicroSpin G-50 column (Pharmacia) equilibrated with STE buffer (50 mM Tris-base, pH 7.5, 1 mM EDTA, 150 mM NaCl), and their final concentration was estimated to be 0.1 pmol μl^{-1}. Subsequently, 15 pmol of ^{32}P-labeled primer and 15 pmol of template (^{32}P-top), or 15 pmol of ^{32}P-labeled template and 15 pmol of primer (^{32}P-bot) were annealed at 75 °C for 2 min in 1× Pol buffer (10 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 1 mM dithiothreitol), followed by slow cooling to room temperature. A primer extension reaction was carried out in the presence of 1 mM dATP, dGTP, dTTP and 70 μCi lyophilized ^{125}I-dCTP (2,200 Ci mmol^{-1}; DuPont-NEN) by 0.5 U exonuclease-free Klenow fragment of DNA polymerase I (US Biochemicals) in a total volume of 50 μl at room temperature. After 10 min the reaction was stopped by addition of 4 μl of 100 mM EDTA, the DNA extracted with phenol:chloroform and the product duplexes purified by MicroSpin G-50 column equilibrated with STE buffer.

DNA–CRP binding. To form the CRP–DNA complex, duplexes (1 pmol) and CRP (33 pmol) were mixed in buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 50 mM KCl, 20 μM cAMP, 50 μg ml^{-1} BSA, 10% glycerol v/v in a total volume of 20 μl and incubated at 25 °C for 30 min. A 10% native PAGE was prerun at 120 V in buffer containing 90 mM Tris-base, pH 8.3, 90 mM boric acid, 1 mM EDTA, 20 μM cAMP for 40 min. Then aliquots of the samples were quickly loaded onto the gel. The electrophoresis was continued for 2.5 h and the gel was fixed in 10% acetic acid (v/v), dried and quantitated with a BAS 1500 Bio-Imaging Analyzer (FUJI).

Analysis of DNA breaks and distances. The intensities of the bands corresponding to the breaks were measured with a BAS 1500 Bio-Imaging Analyzer as described[6]. To calculate the frequency of breaks at a DNA base, the intensity of the corresponding band was divided by the sum of the intensities of all the bands in the lane. Data from three independent experiments were normalized by the percentages of ^{125}I-dC incorporation and of CRP binding. Then the averages and standard deviations (σ) were calculated (Fig. 5*a*,*b*). The difference between distribution of the probabilities of breaks calculated as described[6] and the distribution of the frequencies of breaks was within experimental error, which was estimated as 10%; therefore, only the distribution of the frequencies of breaks is presented. The distances were obtained in the following way. The ^{125}I atom was positioned in cytosine by analogy with the methyl group in thymine, after which the distances from ^{125}I to five heavy atoms of a sugar ring were calculated and then averaged. The free DNA was modeled by a regular duplex averaged over the B-DNA crystal structures[25]. In CRP-bound DNA the ^{125}I–sugar distances were averaged over two cocrystal structures[16,22].

Correspondence should be addressed to I.G.P. *email: igorp@helix.nih.gov*

Received 18 November, 1998; accepted 20 April, 1999.

1. Harrison, S.C. *Nature* **353**, 715–719 (1991).
2. Patikoglou, G. & Burley, S.K. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 289–325 (1997).
3. Steitz, T.A. *Q. Rev. Biophys.* **23**, 205–280 (1990).
4. Travers, A.A. *Annu. Rev. Biochem.* **58**, 427–452 (1989).
5. *Protein–DNA interactions.* (ed. Sauer, R.T.). *Methods Enzymol.* 208 (1991).
6. Panyutin, I.G. & Neumann, R.D. *Nucleic Acids Res.* **25**, 883–887 (1997).
7. Sastry, K.S. *Med. Phys.* **19**, 1361–1370 (1992).
8. Martin, R.F. & Haseltine, W.A. *Science* **213**, 896–898 (1981).
9. Panyutin, I.G. & Neumann, R.D. *Nucleic Acids Res.* **22**, 4979–4982 (1994).
10. Panyutin, I.G. & Neumann, R.D. *Acta Oncol.* **35**, 817–823 (1996).
11. Nikjoo, H., O'Neill, P., Terrissol, M. & Goodhead, D.T. *Int. J. Radiat. Biol.* **66**, 453–457 (1994).
12. Terrissol, M. *Int. J. Radiat. Biol.* **66**, 447–451 (1994).
13. Fried, M.G. & Crothers, D.M. *Nucleic Acids Res.* **11**, 141–158 (1983).
14. Kolb, A., Spassky, A., Chapon, C., Blazy, B. & Buc, H. *Nucleic Acids Res.* **11**, 7833–7852 (1983).
15. Zinkel, S.S. & Crothers, D.M. *Nature* **328**, 178–181 (1987).
16. Schultz, S.C., Shields, G.C. & Steitz, T.A. *Science* **253**, 1001–1007 (1991).
17. Zhurkin, V.B., Lysov, Y.P. & Ivanov, V.I. *Nucleic Acids Res.* **6**, 1081–1096 (1979).
18. Ivanov, V.I. *et al. J. Mol. Biol.* **245**, 228–240 (1995).
19. Garner, M.M. & Revzin, A. *Nucleic Acids Res.* **9**, 3047–3060 (1981).
20. Kim, J., Zwieb, C., Wu, C. & Adhya, S. *Gene* **85**, 15–23 (1989).
21. Maxam, A.M. & Gilbert, W. *Methods Enzymol.* **65**, 499–525 (1980).
22. Parkinson, G., *et al. J. Mol. Biol.* **260**, 395–408 (1996).
23. Lobachevsky, P.N. & Martin, R.F. *Acta Oncol.* **35**, 809–815 (1996).
24. Walicka, M.A., Adelstein, S.J. & Kassis, A.I. *Radiat. Res.* **149**, 142–146 (1998).
25. Gorin, A.A., Zhurkin, V.B. & Olson, W.K. *J. Mol. Biol.* **247**, 34–48 (1995).
26. Ferrin, T.E., Huang, C.C., Jarvis, L.E. & Langridge, R. *J. Mol. Graph.* **6**, 13–27 (1988)

Identifying two ancient enzymes in Archaea using predicted secondary structure alignment

Huimin Xu[1], Rajeev Aurora[2], George D. Rose[3] and Robert H. White[1]

[1]Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0308, USA. [2]Monsanto Company, Mail Zone AA3G, 700 Chesterfield Parkway North, St. Louis, Missouri 63198, USA. [3]Department of Biophysics and Biophysical Chemistry, Johns Hopkins University, School of Medicine, 725 N. Wolfe Street, Baltimore, Maryland 21205, USA.

It is now possible to compare life forms at high levels of detail and completeness due to the increasing availability of whole genomes from all three domains. However, exploration of interesting hypotheses requires the ability to recognize a correspondence between proteins that may since have diverged beyond the threshold of detection by sequence-based methods. Since protein structure is far better conserved than protein sequence, structural information can enhance detection sensitivity, and this is the basis for the field of structural genomics. Demonstrating the effectiveness of this approach, we identify two important but previously elusive Archaeal enzymes: a homolog of dihydropteroate synthase and a thymidylate synthase. The former is especially noteworthy in that no Archaeal homolog of a bacterial folate biosynthetic enzyme has been found to date. Experimental confirmation of the deduced activity of both enzymes is described. Identification of two different proteins was attempted deliberately to help allay concern that predictive success is merely a lucky accident.

Recently, a novel procedure, ORF (an acronym for ostensible recognition of folds), was developed to identify protein homologs in genomic databases[1]. As an initial test, ORF was used to predict a

gene that encodes thymidylate synthase (TS) in the archeon *Methanococcus jannaschii*. At approximately the same time, a second unpublished prediction was made of a gene that encodes an Archaeal homolog of dihydropteroate synthase (DHPS), a folate biosynthetic enzyme. Here, we report experimental confirmation of both predictions.

DHPS and TS are important enzymes, and each is of high interest in its own right. Regarding DHPS, no Archaeal homolog of a bacterial folate biosynthetic enzyme has yet been identified[2–4]. Similarly, a TS enzyme was not apparent among Archaeal genomes, although the protein is highly conserved among other organisms

The ORF procedure[1] is being developed as part of a structural genomics initiative. ORF operates solely on sequence information to predict the secondary structure of both an unknown target protein and all entries in a database of interest and then uses this information in a query-against-all alignment to select likely candidates. Briefly, in ORF a protein's residue sequence is substituted by its predicted secondary structure, calculated using GORII parameters[5] supplemented by helix-capping rules[6]. The resultant string is written in a three-letter 'alphabet': a (α-helix), b (β-strand) and c (coil — that is, all else). Secondary-structure strings that represent both probe and test sequences are aligned in an optimal pair-wise manner using a dynamic programming algorithm[7] and a score is calculated. Often, a length filter is needed to remove small, partially-aligned fragments (for example, a three-helix protein against a probe containing six helices).

In *M. jannaschii*, ORF predicted MJ0301 as a homolog of 7,8-dihydropteroate synthase from bacteria and MJ0757 as a thymidylate synthase. Here, these predictions are confirmed experimentally by expressing each gene in *Escherichia coli* and demonstrating that the gene product does perform the corresponding biosynthetic reaction.

Dihydropteroate synthase

Modified folates function throughout the Archaea[8]. Methanopterin (MPT), the first identified and characterized modified folate[9], functioning as its tetrahydro derivative, serves as a C_1 carrier coenzyme in the methanogenic Archaea[10,11] and as an alternate coenzyme for many reactions previously shown to involve folate. This substitution of cofactors is necessary because methanogens and other Archaea either lack folates altogether or else express them at very low levels[12].

Our current understanding of the biosynthesis of folate[13] and methanopterin[14] (Fig. 1) reveals many similarities. In both molecules, the pterin portions are derived from GTP and the arylamine portions are derived from 4-aminobenzoic acid (*p*AB). Utilizing these molecular precursors, methanopterin and folate are then synthesized in steps having some[15], but not all[16], reactions in common. In both cases, a central and analogous reaction is the condensation of 6-hydroxymethyl-7,8-dihydropterin pyrophosphate with an arylamine to form the main body of the coenzymes. This reaction is catalyzed by dihydrofolate synthase (DHPS) in folate biosynthesis[13]. Despite these many similarities, no Archaeal counterparts have yet been identified for genes encoding bacterial folate biosynthetic enzymes[2–4].

Using ORF[1], DHPS sequences from *Pneumocystis carinii* and *E. coli* were used to probe all open reading frames in the genomic database of *M. jannaschii* (http://www.tigr.org/tdb/mdb/mjdb). ORF identified MJ0301 as the optimal match (after application of a size-filter); the sequence identity[17] between this candidate and the *P. carinii* protein is 19.9%. The rank position of MJ0301 is the best (z-score of 2.18) and second best (z-score of 2.08) for *P. carinii* and *E. coli* probes, respectively. To our knowledge MJ0301 has not been identified previously as an Archaeal homolog of DHPS.

Three-dimensional structures of DHPS from *P. carinii*[18] and *S. aureus*[19] have been determined by X-ray crystallography at 2.0 Å and 2.2 Å resolution, respectively. The enzyme is an eight-stranded α/β barrel. Secondary structure of the *P. carinii* DHPS[18] is predicted with an accuracy of 68% by ORF, though this precise value may

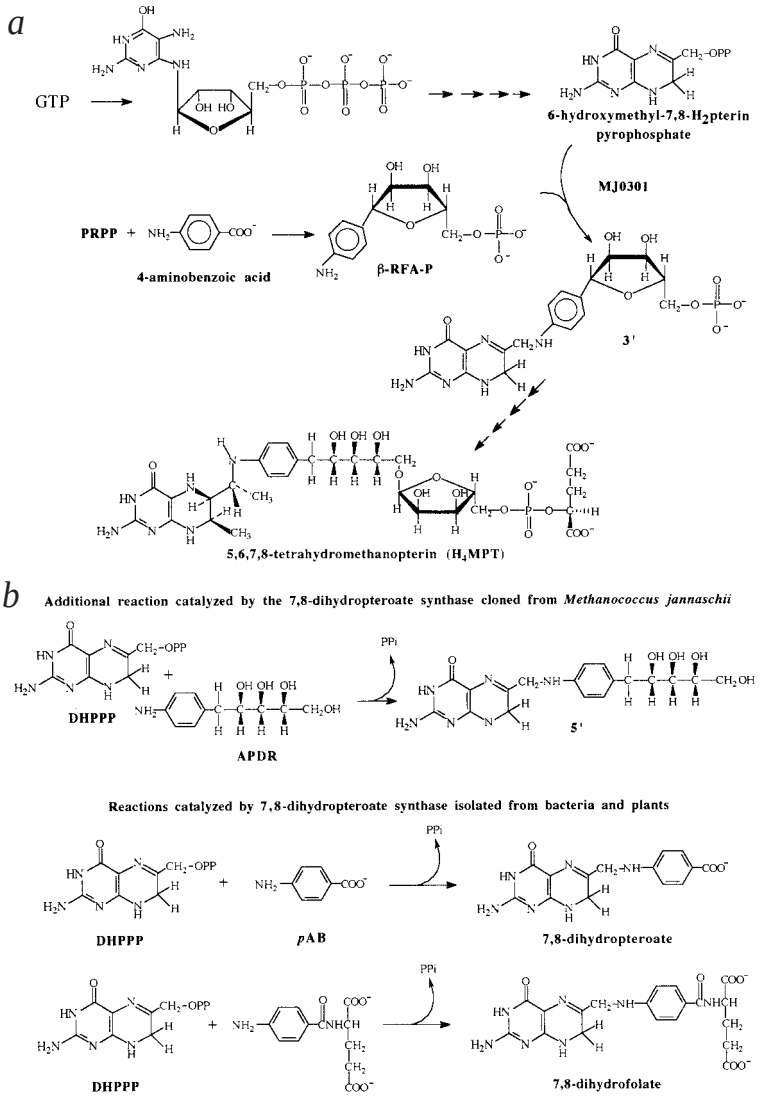

Fig. 1 Biosynthetic pathways. ***a***, Pathway for the biosynthesis of 5,6,7,8-tetrahydromethanopterin. Only the steps in the conversion of 7,8-dihydropterin 3'-phosphate to 6-hydroxymethyl-7,8-dihydropterin pyrophosphate appear to be exactly the same as the reactions involved in H_4-folate biosynthesis[14]. ***b***, Reactions catalyzed by the DHPS cloned from *Methanococcus jannaschii* and the DHPSs isolated or cloned from bacteria and plants.

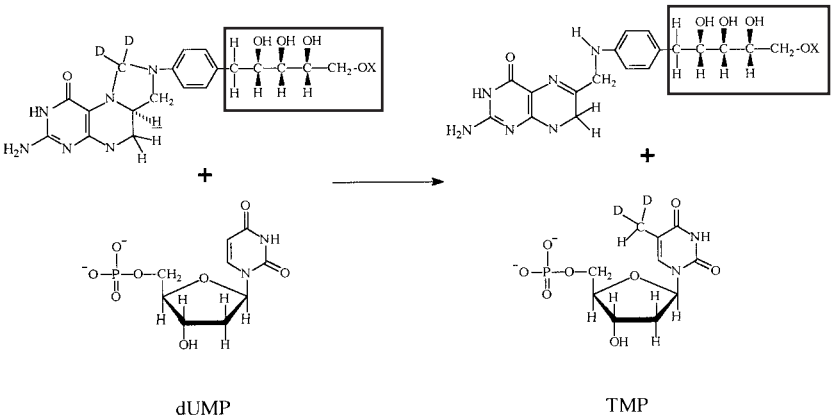

Fig. 2 Reaction proposed to be catalyzed by the thymidylate synthetase from *Methanococcus jannaschii*. The nature of the R group (boxed) is currently unknown. In the actual reaction carried out by the cloned MJ0757 (thymidylate synthetase), [*methylene*-$^2H_2$]-methylenetetrahydrofolate was used as the substrate.

vary slightly depending upon the way helix and strand endpoints are defined. It should be possible to use the known structures[18,19] to guide construction of a three-dimensional model of MJ0301 because coordinates have just been released[20].

Functionally-important residues can be assigned based on either experimental results or homology. Residues involved in catalysis and substrate binding were identified[18,19] in the known DHPS structures; most appear to be conserved in MJ0301, as assessed by either secondary-structure-based (Table 1) or multiple sequence-based alignment with DHPS, with conservative replacements seen at several positions (not shown).

Previously, a different protein, MJ0107, had been identified as a DHPS homolog from sequence-based methods (National Center for Biotechnology database http://www.ncbi.nlm.nih.gov). Using PSI-BLAST[21], we confirmed that MJ0301 is not identified as a DHPS homolog. In particular, the GENPEPT database was searched using the *P. carinii* sequence (gi|2270900) with an E-value cutoff of 1.0×10^{-5}. The search converged after three rounds; MJ0301 was not detected. In the ORF procedure, MJ0107 ranks poorly (114th, z-score = 0.04) and fails to align against many of the catalytic residues in Table 1. Accordingly, we chose to test MJ0301 for DHPS activity. However, MJ0107 should also be tested because it is conceivable that *M. jannaschii* has multiple pteroate synthase genes (see Note added in proof). The MJ0301 protein was expressed in *E. coli*, purified, and pteroate synthase activity[22] was measured, confirming that this protein has the predicted activity.

Yields, reported in Table 2, are based on amounts of coupled products isolated. The best substrate was 4-(β-D-ribofuranosyl)-aminobenzene-5'-phosphate (β-RFA-P) (Fig. 1*a*), the expected natural substrate. Dephosphorylated β-RFA-P was found not to be a substrate, whereas 1-(4-aminophenyl)-1-deoxy-D-ribitol (APDR), a component of the modified folates such as methanopterin, was a substrate. These reactions and their folate biosynthetic analogs are shown in Fig. 1*b*.

The utilization of different substrates in this type of enzymatic reaction is not without precedent: bacterial dihydropteroate synthases can use either *p*AB, *p*ABglutamate, or a series of sulfonamides to form either dihydropteroate, dihydrofolate or a wide range of folate dihydropterin-sulfonamides[23]. The use of APDR and probably other arylamines related to methaniline[15] may represent pathways that can salvage functional substrates from arylamines produced by oxidative (air) cleavage of reduced forms of the coenzymes[24].

Thymidylate synthase

Thymidylate synthase (TS) is one of the most highly conserved enzymes known. Approximately 18% of its residues are absolutely conserved among the 17 known sequences[25]. Yet, surprisingly, the gene for this ubiquitous enzyme was not apparent in Archaeal genomes. Nonetheless, biosynthetic labeling experiments do provide evidence for the existence of TS activity in Archaea[26].

Subsequent to the prediction of MJ0757 as a thymidylate synthase[1], a different candidate (MJ0511) was identified by the TIGR group (see their website http://www.tigr.org/tdb/mdb/mjdb). It is unlikely, though not impossible, that *M. jannaschii* has two genes for the TS function.

An unsuccessful attempt had been made to isolate the Archaeal TS from *Methanobacterium thermoautotrophicum* by following two characteristic, folate-independent activities: tritium exchange of [5-^{3}H]deoxyuridine monophosphate and the dehalogenation of 5-bromodeoxyuridine monophosphate[27]. The protein isolated was probably a deoxyuridylate hydroxymethyltransferase, based in its N-terminal sequence and in later work that identified its full sequence[28]. At first one

Table 1 Equivalent positions in predicted binding sites identified after secondary structure alignment

DHPP Group	Contact type	E. coli DHPS[1]	MJ0301[2]	MJ0107[3]
Phosphate	Hydrogen bond	Asn 22	Asn 24	Val 37
Phosphate	Hydrogen bond	Thr 62	Asn 62	Val 167
Pterin	Hydrogen bond	Asp 96	Glu 115	Asp 224
Pterin	Hydrogen bond	Asn 115	Δ[4]	Val 245
Pterin	Hydrogen bond	Asp 185	Asp 185	Asp 348
Pterin	Hydrogen bond	Lys 221	Lys 227	Glu 438
Pterin	Hydrogen bond	Arg 255	Cys 261	Arg 476
Phosphate	Hydrogen bond	His 257	His 263	Lys 478
Pterin	Hydrophobic	Ile 117	Leu 135	Ala 247
Pterin	Hydrophobic	Phe 190	Leu 193	Ile 365

[1]Contacts observed in the *E. coli* crystal structure DHPS bound to DHPP and assigned by Achari *et al.* (see Fig. 6 in ref. 19).
[2]Residues in MJ0301 that correspond to those in column 3 based on ORF alignment.
[3]Residues in MJ0107 that correspond to those in column 3 based on sequence alignment by ALIGN[34].
[4]Δ indicates that there is an insertion or deletion at this position in the alignment.

Table 2 Products generated during incubation of an *E. coli* extract containing the gene product of MJ0301 with various substrates[1]

Arylamine	Concentration (mM)	Product found or expected	% conversion[2]
β-RFA-P	0.1	7,8-dihydropterin-6-methyl-4-(β-D-ribofuranosyl)-aminobenzene-5'-phosphate (Compound 3' in Fig. 1*a*)	41
APDR	0.1	7,8-dihydropterin-6-methyl-1-(4-aminophenyl)-1-deoxy-D-ribitol (Compound 5' in Fig. 1*b*)	10
β-RFA	0.1	7,8-dihydropterin-6-methyl-4-(β-D-ribofuranosyl)-aminobenzene	nd[3]

[1]Incubations were conducted in the presence of 0.5 mM 6-hydroxymethyl-7,8-dihydropterin pyrophosphate and the indicated concentrations of co-substrates. No products were produced in either control incubations without cell extracts or incubations with cell extracts of BL-21 cells lacking the plasmid.
[2]Percentage of the arylamine converted into product. Number reported is based on the amount recovered.
[3]Not detected.

member of our group adopted a different strategy to identify the enzyme in crude cell extracts of methanogens by assaying the stoichiometric formation of covalent ternary complexes with [6-^{3}H]FdUMP and the 5,10-methylene derivative of either 5,6,7,8-H_4-folate or 5,6,7,8-H_4-methanopterin, but no adduct formation was observed[29].

To test the hypothesis that MJ0757 encodes TS in *M. jannaschii*, the enzyme was overexpressed in *E. coli* and cell extracts were analyzed for the desired activity. It was anticipated that a modified folate would be utilized in this reaction because methanogenic Archaea lack folates, as described above, and coenzyme substitutes have been observed for many folate-type reactions in these cells[8,23]. Accordingly, several modified folates were tested as possible cofactors for the cloned TS (Table 3), but none resulted in formation of thymidine monophosphate (TMP) when incubated with extracts of cloned enzyme, dUMP and labeled formaldehyde.

However, incubation in the presence of H_4-folate did lead to formation of TMP, with the methyl group being derived from labeled formaldehyde as indicated in Fig. 2. This route to the formation of TMP was established by measuring the incorporation of two deuteriums into the thymine generated by acid hydrolysis of the isolated TMP. Both deuterium incorporation and TMP formation occurred only in cell extracts containing the overexpressed enzyme.

Discussion

Folate is not likely to be the natural substrate for Archaeal thymidylate synthase, given that Archaea are typically folate-deficient (halobacteria are an exception). Based on the chemical structures of the modified 5,10-methylene-H_4-folates tested (Table 3), together with the known route of methanopterin biosynthesis[14] (Fig. 1*a*), we conclude that the authentic substrate of MJ0757 is one of the nonmethylated methanopterin biosynthetic intermediates. If so, then the methyl groups in methanopterin enable the cell to segregate energy metabolism from DNA synthesis by utilizing different C_1 carrier molecules: (i) methanopterin, used in energy metabolism and (ii) a non-methylated analog, used in thymine generation for DNA synthesis. Such separation of function would represent yet another example where methylation reactions are deployed to modulate biochemical utilization of specific molecules.

The aligned sequence identity between MJ0301 and other known pteroate synthases is low. Is this a case of divergent or convergent evolution? That is, did these enzymes evolve from a common precursor or have they evolved independently? Given the relatively small number of conceivable protein folds[30], a plausible argument can be made for convergent evolution in this case. If this is true for the other enzymes in methanopterin biosynthesis as well, then these two coenzymes — methanopterin and folate — are end products of entire convergent pathways.

Throughout evolution, viable solutions to metabolic problems have been either recycled or reinvented. From *E. coli* to *Homo sapiens*, similar processes utilize similar molecules and mechanisms. Computational biologists capitalize on this fact with both sequence- and structure-based methods that infer structure and function of unknown proteins from known homologs[31]. At present, sequence-based methods are better developed. However, over an evolutionary time-frame, proteins can diverge to the point that homologs are no longer apparent from sequence alone[31]. Sensitivity can be enhanced by recognizing that not all residues are of equal importance. In particular, an enzyme's function is realized through the chemistry of its active site, which is typically com-

Table 3 Deuterated formaldehyde incorporated into TMP from dUMP by *E. coli* cell extracts with or without overexpressed MJ0757

Cofactors[1]	Cell extract[2]	Atom % of thymine with $^2H_2$ (ftn. 3)	Thymine isolated[4] (nmol)
Tetrahydrofolate	+	95	70
Tetrahydrofolate	–	<1%	2.0
None	+	<1%	2.0
Tetrahydrosarcinapterin	+	<1%	2.0
Tetrahydrosulfofolate	+	<1%	2.0
[*methyl*-^{3}H]-methyl-tetrahydrosarcinapterin	+	<0.00003 (ftn. 5)	nd[6]

[1]Details of the experimental conditions and cofactors used are provided in the Methods.
[2]A '+' or a '–' indicates whether cells used to prepare the cell extract included or excluded protein product derived from MJ0757, respectively.
[3]Atom % is defined as the mole fraction of molecules containing $^2H_2$.
[4]TMP was isolated from the reaction mixture by ion-exchange chromatography and hydrolyzed to thymine, then further purified by chromatography on a C_{18} column and by preparative TLC. Quantitation was based on absorbance at 260 nm of isolated thymine samples, and incorporation of deuterium was measured by GC-MS of the ditrimethylsilyl derivatives of isolated thymine[26].
[5]Based on observed incorporation of ^{14}C.
[6]Not determined.

prised of only a small set of residues embedded in a three-dimensional structure. Inherently, this design lends itself to a mutable sequence within a persisting structural framework. Structure-based fold recognition has exploited this design, with conspicuous success when the homolog is of known structure[32], but with limited success when it is not. However, the examples in this paper offer encouragement that successful, predictive structure-based fold recognition is on the horizon.

Note added in proof: Since acceptance of this paper, the MJ0107 gene, which was identified as a dihydropteroic acid synthase by sequence-based methods, has now been cloned and the corresponding protein overexpressed in E. coli, using the methods described in the text. The resulting soluble protein did not catalyze the condensation of 6-hydroxymethyl-7, 8-dihydropterin and 4-aminibezoic acid to form dihydropteroic acid. Thus, this gene does not encode a dihydropteroate synthase.

Methods

Dihydropteroate synthase activity measurement. MJ0301 was cloned into a pT7-7 expression vector, the desired protein was overexpressed in *E. coli*. The plasmid construct AMJE067 containing *M. jannaschii* gene MJ0301 and the pUC18 vector were obtained from TIGR/ATCC microbial genome special collection. The oligonucleotide primers used to direct the polymerase chain reaction of the MJ0301-containing gene cartridge had sequences: '5-CATGCATATGTTAATGAA-CATTGGAAAA-3' and '5-GATCGGATCCTTAAAATTTTATAATCTTTCC-3'. High level expression of the MJ0301 gene product in the *E. coli* host strain BL21(DE3) was accomplished by constructing a gene cartridge *in vitro* and cloning it into a pT7-7 plasmid such that gene expression is controlled by the T7 phage transcriptional and translational regulatory elements, which are regulated, in turn, by the *lac* control elements. The recombinant plasmid was transformed into the host strain and the cells were grown in LB medium supplemented with 100 μg ml^{-1} ampicillin at 30 °C to an absorbance at 600 nm of 1.0. Protein production was then induced by the addition of isopropylthio-β-D-galactoside (IPTG) to a final concentration of 1 mM. After growth of the cells in the presence of IPTG for four hours, the cells were harvested by centrifugation (4,000 × g for 5 min), and frozen at -20 °C until used. High expression of the desired protein was confirmed by SDS-PAGE (12% polyacrylamide) of the SDS-soluble cellular proteins. Presence of the desired enzyme was also confirmed by measuring its corresponding activity.

Enzymatic activity was measured in cell extracts obtained by sonication[15]. Cell pellets (~200 mg wet weight) were suspended in 2 ml of 50 mM TES, 10 mM $MgCl_2$ and 20 mM mercaptoethanol, pH 7.0, sonicated for 3 min at 0 °C and centrifuged for 10 min at 16,000 × g. The resulting crude extracts (0.4 ml, ~10 mg ml^{-1} protein) were incubated under anaerobic conditions in argon at 50 °C for 2 h with 6-hydroxymethyl-7,8-dihydropterin pyrophosphate and various acceptor arylamines, as indicated in Table 2. No products were produced in either control incubations without cell extracts or incubations with cell extracts of BL-21 cells lacking the plasmid. Samples incubated with β-RFA-P were adjusted to pH 9–10 with 1.0 M NaOH after addition of 100 μl of 0.1 M glycine buffer, pH 10.4, containing 1 mM $ZnCl_2$ and 1 mM $MgCl_2$, and incubated for 2 h with 0.3 units of *E. coli* alkaline phosphatase at 37°, and then adjusted to pH 7.0. Next, 200 μl of a solution (0.02 M iodine and 0.06 M KI) were mixed with each sample to oxidize the dihydropterin to pterin. The iodine color was removed after 2 min at room temperature by titration with a saturated solution of sodium bisulfite. After addition of 50 μl of 1 M sodium acetate buffer, pH 5.5, samples were placed on a C_{18} reverse phase column (0.6 × 4.5 cm) and eluted first with distilled water (3 ml), then with a step gradient consisting of two 2.4 ml portions of methanol in water. The methanol concentrations used were increased in consecutive steps from 5% to 10%, 20%, and 40% . Components in the fractions (2.4 mL) were measured by fluorescence and UV-visible absorbance to establish the elution position of the compounds. These elution conditions allowed for the complete separation of the products of the reactions, which eluted with the early fractions, from the reactants that eluted with the 40% methanol. Initial identification of the enzymatic products was established by their UV-visible spectrum and their elution position from the C_{18} column when compared to known synthetic compounds. Each C_{18}- purified product was then subjected to Zn/HCl reductive cleavage to 6-methylpterin, and its respective arylamine product was characterized as described[22].

Thymidylate synthase activity measurement. Experimental details of overexpression were identical to those described for DHPS above, using plasmid construct AMJGV30 containing *M. jannaschii* gene MJ00757 and oligonucleotide primers: 5'-CATGCATATGGTATTAAACATCAATTCT-3' and 5'-GATCGGATCCTTATTTTTTAGGGCTTTTC-3'.

Each thymidylate synthase activity assay (Table 3) was conducted with 0.4 ml of cell extract (~10 mg ml^{-1} protein in 50 mM TES, 10 mM Mg^{2+}, pH 7.0, buffer containing 20 mM mercaptoethanol) to which anaerobic solutions of individual reactants were added. Based on volumes used in each experiment, the following concentrations of substrates were present during the incubations: tetrahydrofolate experiments: 5.5 mM of $^2H_2$-formaldehyde, 2.2 mM dUMP and 0.55 mM tetrahydrofolate; tetrahydrosulfofolate fragment experiments: 53.9 mM of $^2H_2$-formaldehyde, 1.6 mM dUMP and 0.14 mM tetrahydrosulfofolate fragment; and tetrahydrosarcinapterin experiments: 4.7 mM of $^2H_2$-formaldehyde, 1.9 mM dUMP and 0.71 mM tetrahydrosarcinapterin. Samples were incubated under argon at 50 °C for 2 h. The tetrahydrosulfofolate fragment is the 5,6,7,8-tetrahydro derivative of pterin-6-methyl-1-(4-aminophenyl)1-deoxy-D-ribitol, an intermediate in the biosynthesis of methanopterin. The compound is a hydrolytic fragment of the modified folate found in *Sulfolobus solfataricus*[33] and was synthesized as described[14].

Acknowledgments

Supported by a National Science Foundation grant to R.H.W. and a National Institutes of Health grant to G.D.R. We thank M.E. Rasche for running the experiments to trap the covalent complex between FdUMP and thymidylate synthase, K. Harich for preforming the GC-MS analyses, and D. Grahame for supplying the [*methyl*-^{3}H]-methyl-tetrahydrosarcinapterin.

Correspondence should be addressed to R.H.W. *email: rhwhite@vt.edu*

Received 18 November, 1998; accepted 13 May, 1999.

1. Aurora, R. & Rose, G.D. *Proc. Natl. Acad. Sci. USA* **95**, 2818–2823 (1998).
2. Bult, C.J. *et al. Science* **273**, 1058–1073 (1997).
3. Smith, D.R. *et al. J. Bacteriol.* **179**,7135–7155 (1997).
4. Klenk, H.P. *et al. Nature* **390**, 364–370 (1997).
5. Garnier, J. & Robson, B. In *The Development of the prediction of protein structure* (ed. Fasman, G.) 417–465 (Plenum, New York; 1989).
6. Aurora, R. & Rose, G.D. *Protein Sci.* **7**, 21–38 (1998).
7. Needleman, S.B. & Wunsch, C.D. *J. Mol. Biol.* **48**, 443–453 (1970).
8. White, R.H. *Biochemistry* **32**, 745–753 (1993).
9. Keltjens, J.T., Huberts, M.J., Laarhoven, W.H. & Vogels, G.D. *Eur. J. Biochem.* **130**, 537–544 (1983).
10. DiMarco, A.A., Bobik, T.A. & Wolfe, R.S. *Annu. Rev. Biochem.* **59**, 355–394 (1990).
11. Keltjens, J.T., Raemakers-Franken, P.C. & Vogels, G.D. In *Microb. Growth C$_1$ Compd.* [7th Int. Sym.] 135–150 (Intercept, Andover, UK; 1993).
12. Leigh, J.A. *Appl. Environ. Microbiol.* **45**, 800–803 (1983).
13. Green, J.M., Nichols, B.P. & Matthews, R.G. In *Escherichia coli and Salmonella, Cellular and Molecular biology*. (ed. Neidhardt, F.C.) 665–673 (American Society for Microbiology; 1996).
14. White, R.H. *Biochemistry* **35**, 3447–3456 (1996).
15. White, R.H. *Biochemistry* **29**, 5397–5404 (1990).
16. Howell, D.M. & White, R.H. *J. Bacteriol.* **179**, 5165–5170 (1997).
17. Myers, E.W. & Miller, W. *Bull. Math. Biol.* **51**, 5–37 (1989).
18. Achari, A. *et al. Nature Struct. Biol.* **4**, 490–497 (1997).
19. Hampele, I.C. *et al. J. Mol. Biol.* **268**, 21–30 (1997).
20. Bernstein, F.C. *et al. J. Mol. Biol.* **112**, 535–542 (1977).
21. Altschul, S.F. *et al. Nucleic Acids Res.* **25**, 3389–3402 (1997).
22. White, R.H. *Methods Enzymol.* **44**, 391–401 (1997)
23. Shiota, T. In *Folates and pterins, Vol. 1, chemistry and biochemistry of folates* (eds Blakley, R.L. & Benkovic, S.J.) 121 (John Wiley & Sons, New York; 1984).
24. White, R.H. *J. Bacteriol.* **162**, 516–520 (1985).
25. Perry, K.M. *et al. Proteins*, **8**, 315–333 (1990).
26. Nyce, G.W. & White, R.H. *J. Bacteriol.* **178**, 914–916 (1996).
27. Krone, U.E., McFarlan, S.C. & Hognekamp, H.P.C. *Eur. J. Biochem.* **220**, 789–794 (1994).
28. Vaupel, M., Dietz, H., Linder, D. & Thauer, R.K. *Eur. J. Biochem.* **236**, 294–300 (1996).
29. Priest, D.G. & Doig, M.T. *Methods Enzymol.* **122**,313–319 (1986).
30. Holm, L. & Sander, C. *Science* **273**, 595–603 (1996).
31. Doolittle, R.F. *Annu. Rev. Biochem.* **64**, 287–314 (1995).
32. Gibrat, J., Madej, T. & Bryant, S.H. *Curr. Opin. Struct. Biol.* **6,** 377–385 (1996).
33. Zhou, D. & White, R.H. *J. Bacteriol.* **174**, 4576–4582 (1992).
34. Myers, E and Miller W. *Comput. Appl. Biosci.* **4,** 11–17 (1988).