Science
AAAS

Transcribing RNA Polymerase II Is Phosphorylated at CTD Residue Serine-7

Rob D. Chapman *et al.*

Science **318**, 1780 (2007);
DOI: 10.1126/science.1145977

Transcribing RNA Polymerase II Is Phosphorylated at CTD Residue Serine-7

Rob D. Chapman,[1] Martin Heidemann,[1] Thomas K. Albert,[2]* Reinhard Mailhammer,[1] Andrew Flatley,[2] Michael Meisterernst,[2]* Elisabeth Kremmer,[2] Dirk Eick[1]†

RNA polymerase II is distinguished by its large carboxyl-terminal repeat domain (CTD), composed of repeats of the consensus heptapeptide Tyr[1]-Ser[2]-Pro[3]-Thr[4]-Ser[5]-Pro[6]-Ser[7]. Differential phosphorylation of serine-2 and serine-5 at the 5′ and 3′ regions of genes appears to coordinate the localization of transcription and RNA processing factors to the elongating polymerase complex. Using monoclonal antibodies, we reveal serine-7 phosphorylation on transcribed genes. This position does not appear to be phosphorylated in CTDs of less than 20 consensus repeats. The position of repeats where serine-7 is substituted influenced the appearance of distinct phosphorylated forms, suggesting functional differences between CTD regions. Our results indicate that restriction of serine-7 epitopes to the Linker-proximal region limits CTD phosphorylation patterns and is a requirement for optimal gene expression.

Differential phosphorylation of CTD residues of the large subunit of eukaryotic RNA polymerase II (Pol II) occurs during the transcription cycle and appears to orchestrate the recruitment, activation, and displacement of various factors involved in transcription and mRNA processing (*1*, *2*). A variety of kinases have been identified, with phosphorylation activity directed toward the amino acids tyrosine-1 (Abl1/2), serine-2 (CTDK1, CDK9, and DNA-PK), serine-5 (ERK1/2 and CDK7-9), and serine-7 (DNA-PK) (*3*). The mammalian CTD is >99% conserved across species and possesses almost double the length of its yeast counterparts (*4*). A minimum length of CTD is required to support the growth of yeast or mammalian cells. However, this is dependent on the number and position of consensus and nonconsensus repeats, which suggests that CTD function is composed of both sequence and length (*5*–*8*). Of the 52 mammalian CTD repeats, 21 obey the consensus sequence and lie largely proximal to the Linker region (fig. S1). The distal C-terminal region deviates from this consensus, predominantly at position 7. These nonconsensus repeats may affect the binding of specific factors or may serve to prevent phosphorylation at the position of deviation. Indeed, studies in vivo suggest that they are equivalent to consensus repeats for functions such as splicing of the fibronectin extra domain I exon (*9*) but not for maintenance of long-term cell viability (*5*, *6*).

To investigate the role of the CTD repeat structure on its phosphorylation, we have established a system that allows the comparison of CTDs of different lengths and repeat compositions in vivo. Recombinant polymerases are engineered with a point mutation conferring resistance to α-amanitin, allowing the endogenous polymerase to be inhibited (and degraded) after addition of α-amanitin but without affecting recombinant polymerase activity (*5*, *10*). Monoclonal antibodies (mAbs) were produced against the CTD phosphoserine epitopes Ser[2]-P, Ser[5]-P, and Ser[7]-P (Fig. 1A and fig. S2). In preparing these antibodies, we considered earlier findings that

[1]Institute of Clinical Molecular Biology and Tumour Genetics, GSF-Research Center of Environment and Health, Munich Center for Integrated Protein Science (CiPSM), Marchioninistrasse 25, 81377 Munich, Germany. [2]Institute of Molecular Immunology, GSF-Research Center of Environment and Health, Munich Center for Integrated Protein Science, Marchioninistrasse 25, 81377 Munich, Germany.

*Present address: Institute of Tumor Biology, University Muenster, Robert-Koch-Strasse 43, 48149 Muenster, Germany.
†To whom correspondence should be addressed. E-mail: eick@gsf.de

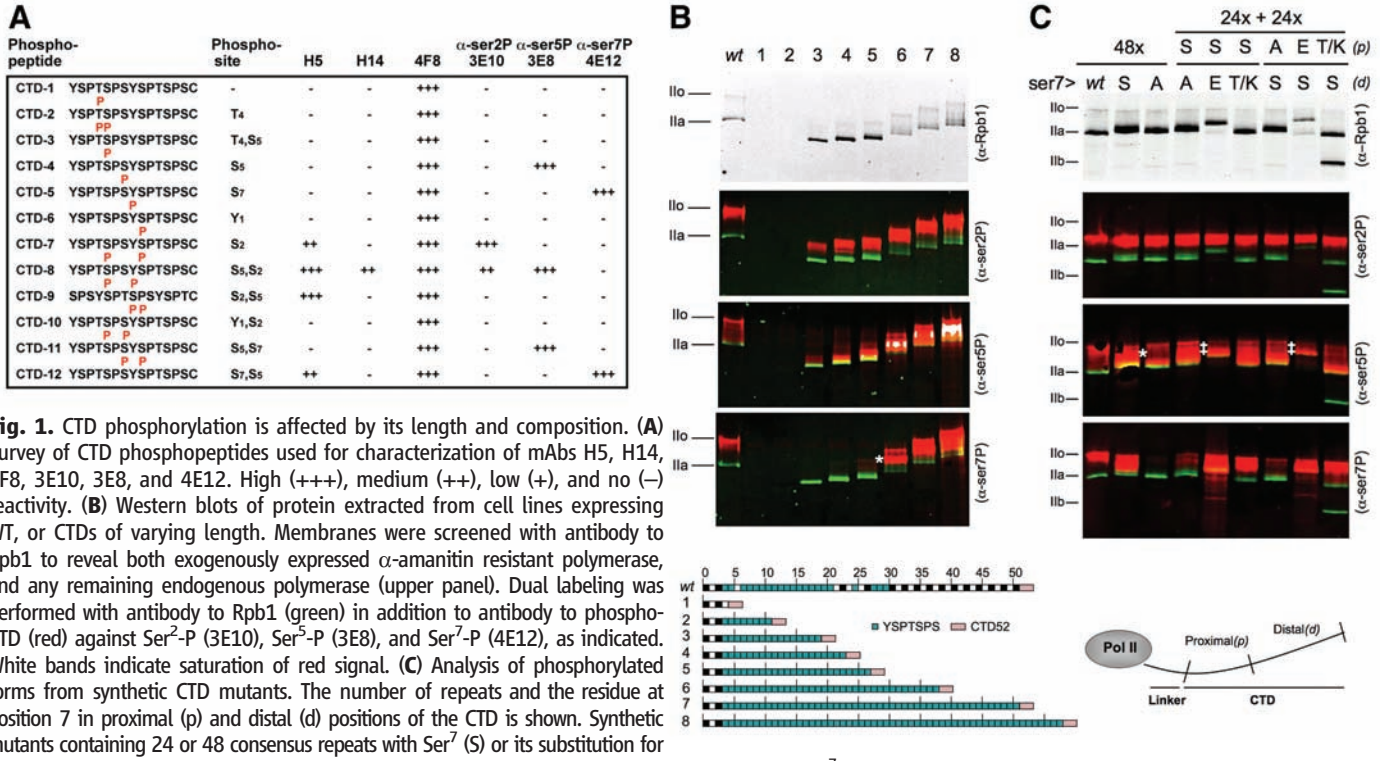


Fig. 1. CTD phosphorylation is affected by its length and composition. (**A**) Survey of CTD phosphopeptides used for characterization of mAbs H5, H14, 4F8, 3E10, 3E8, and 4E12. High (+++), medium (++), low (+), and no (–) reactivity. (**B**) Western blots of protein extracted from cell lines expressing WT, or CTDs of varying length. Membranes were screened with antibody to Rpb1 to reveal both exogenously expressed α-amanitin resistant polymerase, and any remaining endogenous polymerase (upper panel). Dual labeling was performed with antibody to Rpb1 (green) in addition to antibody to phospho-CTD (red) against Ser[2]-P (3E10), Ser[5]-P (3E8), and Ser[7]-P (4E12), as indicated. White bands indicate saturation of red signal. (**C**) Analysis of phosphorylated forms from synthetic CTD mutants. The number of repeats and the residue at position 7 in proximal (p) and distal (d) positions of the CTD is shown. Synthetic mutants containing 24 or 48 consensus repeats with Ser[7] (S) or its substitution for alanine (A) confirm the specificity of the 4E12 mAb. Chimeras of consensus repeats (lanes labeled S) and repeats with Ser[7] mutated to alanine (A), glutamic acid (E), and threonine/lysine (T/K) affect the appearance of a Ser[5]-P/Ser[7]-P reactive band when Ser[7] is absent in the Linker proximal region. Mutants containing T/K-rich sequences at position 7, proximal to the Linker, result in degradation of polymerase to the CTD-less, Pol IIb form.

showed that the functional unit of the CTD is not the heptad repeat itself but is in a sequence lying within heptapeptide pairs (*11*, *12*). Thus, in the production and testing of our antibodies, a panel of di-heptapeptides with various modifications was used. Analysis of our antibodies and commercially available antibodies revealed that some recognition profiles were limited by modifications on neighboring repeats. For example, the α-Ser^7-P antibody (4E12) is affected by upstream, but not downstream, Ser^5-P (Fig. 1A).

Combining these tools, we compared the phosphorylation of wild-type (WT) CTD with that of different lengths of consensus repeats (Fig. 1B). If all repeats are equally accessible to CTD kinases, we should expect intensities of phosphorylation signals for WT and mutants 1 to 8 proportional to CTD length. Dual labeling of membranes with α-Rpb1 antibody (green signal) (mAb Pol3/3 recognizes an epitope outside the CTD) and with α-phospho-CTD antibody (red signal) reveals forms of different mobility—the rapidly migrating, unmodified IIa form and the slower, modified IIo form. For mutants containing 16 to 24 consensus repeats (mutants 3 to 5), the majority of Pol II is not efficiently phosphorylated and accumulates in the IIa form (green). Mutants 1 and 2 are no longer visible at this time point because they are unable to support their own expression. Within the IIo form, Ser^2-P appears in a sharp, slow migrating band, whereas in longer CTDs (mutants 6 to 8), Ser^5-P appears largely in a band (white saturation) migrating between the Ser^2-P band and IIa, which suggests that at least two populations of phosphorylated CTD exist in vivo at any time: Ser^2-P alone and Ser^5-P alone. These data are supported by both the recognition profiles of the antibodies and previous work showing a shift in IIo to a faster migrating form upon treatment with a Ser^2-kinase inhibitor (*13*). Antibody raised against Ser^7-P revealed the existence of this epitope in vivo, which is distributed among the major Ser^2-P and Ser^5-P reactive bands. The epitope is lacking from the Ser^5-P band that appears just above the IIa form. Strong reactivity of α-Ser^7-P is detectable for a band between IIa and IIo (Fig. 1B, *). Furthermore, although Ser^2-P and Ser^5-P appear in all mutants, Ser^7-P appears only in mutants with more than 24 repeats (mutants 4 to 8).

To investigate the effect of nonconsensus repeats on the distribution of phosphorylation, a panel of CTD mutants (fig. S3) was analyzed for their reactivity against phospho-CTD antibodies (Fig. 1C). α-Ser^7-P does not recognize a mutant lacking Ser^7 (48xS7A) but strongly recognizes mutants containing Ser^7 substituted with glutamic acid (S7E), indicating either that this antibody recognizes a CTD conformation or that S7E can structurally mimic Ser^7-P for antibody recognition. Furthermore, replacement of Ser^7 with alanine prevents recognition of the intermediate band between IIa and IIo (Fig. 1C, *) by α-Ser^5-P, suggesting that this form may be Ser^7-P–dependent.

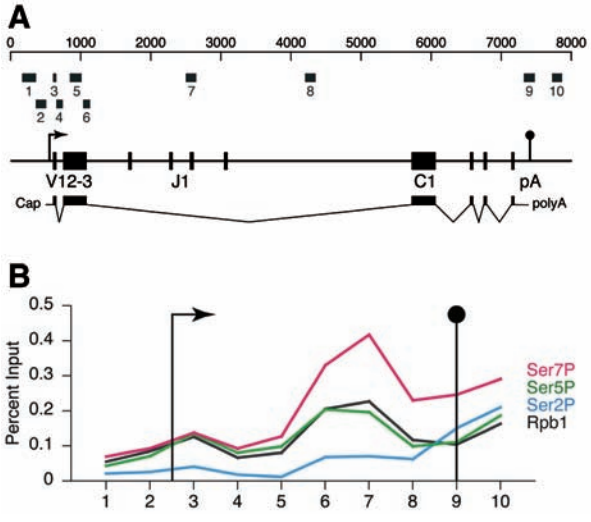
Because deviations from serine at position 7 in the WT CTD are concentrated in its distal region, chimeras were produced to assess the effect of proximal and distal positioning of nonconsensus repeats. The two chimeras of consensus repeats, and repeats containing S7E substitutions, produce a form that migrates between IIa and IIo (Fig. 1C, ‡). The proximal positioning of nonconsensus repeats (S7A and S7T/K) affects the appearance of a form similar in mobility to the intermediate IIo Ser$^{5/7}$-P–reactive band (Fig. 1, B and C, *) seen in mutants of >35 pure consensus repeats. (Fig. 1, B and C; for more detail, see fig. S4).

To determine whether Ser^7 phosphorylation is a physiological event during the transcription cycle, chromatin immunoprecipitation (ChIP) experiments were conducted. A detailed example is shown for the T cell receptor beta (TCRβ) gene locus (Fig. 2A). Ser^7 was phosphorylated on transcribing Pol II, appearing strongly at the promoter and increasing toward the 3′ region of TCRβ (Fig. 2B). The differences in Ser^2 phosphorylation that we observe, compared with earlier data, may result from the antibodies used, because the H5 antibody preferentially recognizes repeats with phosphorylated Ser^2 and Ser^5 (Fig. 1A).

Given that Ser^7 is phosphorylated across TCRβ and all other genes tested (*GAPDH*, *RPLPO*, and *RPS27*), the ability of synthetic polymerases to transcribe and produce mature mRNA from the *c-myc* (Fig. 2C) and *pes1* genes (fig. S5) was analyzed. The effect on *c-myc* and *pes1* mRNA levels of Ser^7 substitution to E or K/T appears dependent on its position, either proximal or distal to the Linker, suggesting again that functional differences exist between these regions. Substitution of Ser^7 to the non-phosphoacceptor, alanine, did not obviously affect mRNA levels, nor did it affect the long-term growth of cell lines, although viability was compromised (fig. S6). This may be due to the effect of this mutation on small nuclear RNA genes (*14*).

ChIP experiments revealed that S7E-containing mutants do not stably associate with any of the genes tested, providing an explanation for the deficit in mRNA observed for mutants containing S7E in the Linker-proximal region. Mutants containing either 48 consensus or S7A repeats appear to be recruited to protein coding genes at similar levels (*14*).

We conclude that the nature of the amino acid at position 7 of the Linker-proximal CTD region is important in steps involved in the stable association of Pol II with class II genes. Accumulation of Ser^7-P in the 3′ region of the TCRβ gene may suggest a role in transcription and/or 3′ RNA processing of some protein-coding genes. We are now able to expand previous models for the cycle of CTD modification across genes that are transcribed by RNA polymerase II (*15*), not only to show how potential phosphorylation patterns change from 5′ to 3′ regions across a gene but also to speculate as to the region of the CTD in which they occur. Phosphorylation of Ser^7 in the proximal part of CTD and replacement of Ser^7 by other amino



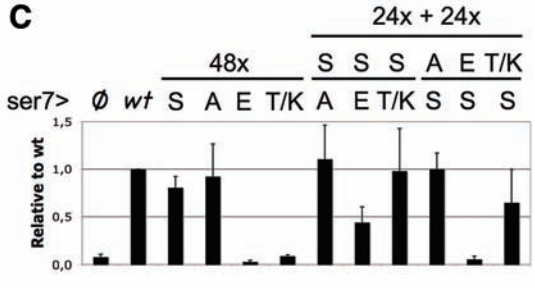


Fig. 2. ChIP analysis of CTD modifications. (**A**) Scale drawing of the rearranged TCRβ gene, with black boxes indicating exons of the variable segment V12-3, of the joining region cluster J1, and of the constant region C1. Gray boxes on top show relative location of ChIP quantitative real-time polymerase chain reaction (QPCR) amplicons 1 to 10. (**B**) QPCR analysis of ChIP analyses in Jurkat cells with antibodies against Rpb1 and indicated phospho-CTD antibodies. The relative location of the transcriptional start site (arrow) and pA (lollipop) are indicated. (**C**) Transcriptional activity of CTD mutants. QPCR of *c-myc* mRNA using primers specific for the mature, spliced form. To control for the effects of α-amanitin treatment, values were normalized against a *7SK* RNA control for the same sample. Myc expression is shown relative to WT levels. Labeling as in Fig. 1B; Ø, without recombinant Pol II. Error bars, SD.

acids in the distal part of CTD may constitute an added layer of gene regulation by mammalian RNA polymerases.

References and Notes

1. A. Meinhart, T. Kamenski, S. Hoeppner, S. Baumli, P. Cramer, *Genes Dev.* **19**, 1401 (2005).
2. H. P. Phatnani, A. L. Greenleaf, *Genes Dev.* **20**, 2922 (2006).
3. B. Palancade, O. Bensaude, *Eur. J. Biochem.* **270**, 3859 (2003).
4. E. Barron-Casella, J. L. Corden, *J. Mol. Evol.* **35**, 405 (1992).
5. M. S. Bartolomei, N. F. Halden, C. R. Cullen, J. L. Corden, *Mol. Cell. Biol.* **8**, 330 (1988).
6. R. D. Chapman, M. Conrad, D. Eick, *Mol. Cell. Biol.* **25**, 7665 (2005).
7. M. L. West, J. L. Corden, *Genetics* **140**, 1223 (1995).
8. Y. Pei, S. Hausmann, C. K. Ho, B. Schwer, S. Shuman, *J. Biol. Chem.* **276**, 28075 (2001).
9. M. de la Mata, A. R. Kornblihtt, *Nat. Struct. Mol. Biol.* **13**, 973 (2006).
10. M. Meininghaus, R. D. Chapman, M. Horndasch, D. Eick, *J. Biol. Chem.* **275**, 24375 (2000).
11. J. W. Stiller, M. S. Cook, *Eukaryot. Cell* **3**, 735 (2004).
12. E. Vojnic, B. Simon, B. D. Strahl, M. Sattler, P. Cramer, *J. Biol. Chem.* **281**, 13 (2006).
13. J. Medlin *et al.*, *EMBO J.* **24**, 4154 (2005).
14. S. Egloff *et al.*, *Science* **318**, 1777 (2007).
15. S. Buratowski, *Nat. Struct. Biol.* **10**, 679 (2003).

16. Supported by grants from the Deutsche Forschungsgemeinschaft (SFB/TR5 and SFB684), Bundesministerium für Bildung und Forschung (QuantPRO, 0313860), Boehringer Ingelheim Fonds, and Fonds der Chemischen Industrie. We thank A. Meinhart and P. Cramer for critical comments on this manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/318/5857/1780/DC1
Materials and Methods
Figs. S1 to S6
References

4 June 2007; accepted 30 October 2007
10.1126/science.1145977

A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea

Ivan A. Berg,[1] Daniel Kockelkorn,[1] Wolfgang Buckel,[2] Georg Fuchs[1]*

The assimilation of carbon dioxide (CO_2) into organic material is quantitatively the most important biosynthetic process. We discovered that an autotrophic member of the archaeal order Sulfolobales, *Metallosphaera sedula*, fixed CO_2 with acetyl–coenzyme A (acetyl-CoA)/propionyl-CoA carboxylase as the key carboxylating enzyme. In this system, one acetyl-CoA and two bicarbonate molecules were reductively converted via 3-hydroxypropionate to succinyl-CoA. This intermediate was reduced to 4-hydroxybutyrate and converted into two acetyl-CoA molecules via 4-hydroxybutyryl-CoA dehydratase. The key genes of this pathway were found not only in *Metallosphaera* but also in *Sulfolobus*, *Archaeoglobus*, and *Cenarchaeum* species. Moreover, the Global Ocean Sampling database contains half as many 4-hydroxybutyryl-CoA dehydratase sequences as compared with those found for another key photosynthetic CO_2-fixing enzyme, ribulose-1,5-bisphosphate carboxylase-oxygenase. This indicates the importance of this enzyme in global carbon cycling.

The dominant extant autotrophic pathway is the Calvin-Bassham-Benson cycle, which may have evolved late in evolution (*1*). Three additional autotrophic pathways are known: the reductive citric acid cycle (Arnon-Buchanan cycle), the reductive acetyl-CoA pathway (Wood-Ljungdahl pathway), and the 3-hydroxypropionate cycle. These pathways differ in several ways [e.g., with respect to energy demand, available reducing compounds, requirement for metals (Fe, Co, Ni, and Mo), usage of coenzymes, and oxygen sensitivity of enzymes]. These criteria determine the distribution of the pathways in autotrophic organisms in different habitats, in addition to their phylogeny.

In the thermoacidophilic Archaea *Sulfolobus*, *Acidianus*, and *Metallosphaera* spp., CO_2 fixation appears to proceed via a modified, but until now unresolved, variant of the 3-hydroxypropionate cycle (*2*–*6*) (Fig. 1). The 3-hydroxypropionate cycle was originally discovered in the phototrophic bacterium *Chloroflexus aurantiacus* (*7*, *8*); it involves the carboxylation of acetyl-CoA to malonyl-CoA by the biotin-dependent acetyl-CoA carboxylase. Because Archaea contain only trace amounts of fatty acids, if any, in their lipids, acetyl-CoA carboxylase cannot serve as the key enzyme of fatty acid synthesis. It seems that this enzyme functions in another metabolic pathway. Malonyl-CoA is reduced via malonate semialdehyde to 3-hydroxypropionate, which is further reductively converted to propionyl-CoA. Propionyl-CoA is carboxylated to (*S*)-methylmalonyl-CoA by a carboxylase, which is similar or identical to acetyl-CoA carboxylase. (*S*)-Methylmalonyl-CoA is isomerized to (*R*)-methylmalonyl-CoA, followed by carbon rearrangement to succinyl-CoA by coenzyme vitamin B_{12}–dependent methylmalonyl-CoA mutase. In *Chloroflexus*, succinyl-CoA is converted to (*S*)-malyl-CoA, which is cleaved by (*S*)-malyl-CoA lyase to acetyl-CoA (thus regenerating the CO_2-acceptor molecule) and glyoxylate (*8*). However, this succinyl-CoA conversion via (*S*)-malyl-CoA was not verified in Sulfolobales (*3*, *4*, *6*), leaving the problem of how acetyl-CoA is regenerated in this group of organisms.

A comparison of the genomes of autotrophic Archaea that have acetyl-CoA carboxylase genes revealed several groups, which differ with respect to the pattern of other characteristic genes and enzymes (table S1). One group comprising autotrophic members of the genera *Metallosphaera*, *Sulfolobus*, *Archaeoglobus*, and *Cenarchaeum* contains the genes for acetyl-CoA/propionyl-CoA carboxylase and methylmalonyl-CoA mutase. These organisms differ considerably in their energy metabolism and habitats. *Metallosphaera* and *Sulfolobus* spp. metabolize sulfur, pyrite, or hydrogen in volcanic areas (*9*); *Archaeoglobus* spp. are anaerobic sulfate reducers (*10*); and *Cenarchaeum* sp. is a symbiont of marine animals (*11*), which is closely related to *Nitrosopumilus* sp., a marine ammonia oxidizer (*12*). All these organisms contain the gene for 4-hydroxybutyryl-CoA dehydratase, a [4Fe–4S] cluster and flavin adenine dinucleotide–containing enzyme, which catalyzes the elimination of water from 4-hydroxybutyryl-CoA by a ketyl radical mechanism yielding crotonyl-CoA (*13*, *14*). This reaction is known to play a role in a few anaerobic bacteria such as *Clostridium aminobutyricum*, which ferment 4-aminobutyrate via 4-hydroxybutyryl-CoA (*15*).

We suspected that this group of Archaea is able to reduce succinyl-CoA to 4-hydroxybutyrate and convert it with the participation of 4-hydroxybutyryl-CoA dehydratase into two molecules of acetyl-CoA. Cell extracts of *M. sedula*, indeed, catalyzed a rapid (500 nmol min^{-1} mg^{-1} protein at 65°C) reduction of succinyl-CoA via succinate semialdehyde to 4-hydroxybutyrate with NADPH, the reduced form of nicotinamide adenine dinucleotide phosphate ($NADP^+$) (*16*). Extracts also converted 4-hydroxy[1-^{14}C]butyrate to [^{14}C]acetyl-CoA at 50 nmol min^{-1} mg^{-1} protein at 65°C in the presence of Mg–adenosine triphosphate (MgATP), CoA, and nicotinamide adenine dinucleotide (NAD^+) (Fig. 2A). High-performance liquid chromatography (HPLC) analysis of the reaction course revealed labeled 4-hydroxybutyryl-CoA, crotonyl-CoA, and 3-hydroxybutyryl-CoA as intermediates (Fig. 2B). A plot of the relative amounts of radioactivity in the individual products versus time indicated that the order of intermediates was as expected and that acetyl-CoA was the end product (Fig. 2C). The organism grows at 75°C in a $H_2/O_2/CO_2$ mixture with a minimum generation time of 8 hours, which corresponds to a specific carbon fixation rate in vivo of 120 nmol min^{-1} mg^{-1} protein (*17*). Because two molecules of carbon are fixed in the proposed pathway, the minimal in

[1]Mikrobiologie, Fakultät Biologie, Universität Freiburg, Schänzlestraße 1, D-79104 Freiburg, Germany. [2]Mikrobiologie, Fachbereich Biologie, Universität Marburg, D-35032 Marburg, Germany.

*To whom correspondence should be addressed. E-mail: georg.fuchs@biologie.uni-freiburg.de