Comparative Analysis of Folding and Substrate Binding Sites between Regulated Hexameric Type II Citrate Synthases and Unregulated Dimeric Type I Enzymes[†,‡]

Nham T. Nguyen,[§,||] Robert Maurus,[§,||] David J. Stokell,[⊥] Ayeda Ayed,[⊥] Harry W. Duckworth,[⊥] and Gary D. Brayer*[,§]

Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver V6T 1Z3, Canada, and Department of Chemistry, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

Received February 27, 2001; Revised Manuscript Received July 2, 2001

ABSTRACT: We describe the first structure determination of a type II citrate synthase, an enzyme uniquely found in Gram-negative bacteria. Such enzymes are hexameric and are strongly and specifically inhibited by NADH through an allosteric mechanism. This is in contrast to the widespread dimeric type I citrate synthases found in other organisms, which do not show allosteric properties. Our structure of the hexameric type II citrate synthase from *Escherichia coli* is composed of three identical dimer units arranged about a central 3-fold axis. The interactions that lead to hexamer formation are concentrated in a relatively small region composed of helix F, FG and IJ helical turns, and a seven-residue loop between helices J and K. This latter loop is present only in type II citrate synthase sequences. Running through the middle of the hexamer complex, and along the 3-fold axis relating dimer units, is a remarkable pore lined with 18 cationic residues and an associated hydrogen-bonded network. Also unexpected was the observation of a novel N-terminal domain, formed by the collective interactions of the first 52 residues from the two subunits of each dimer. The domain formed is rich in β-sheet structure and has no counterpart in previous structural studies of type I citrate synthases. This domain is located well away from the dimer−dimer contacts that form the hexamer, and it is not involved in hexamer formation. Another surprising observation from the structure of type II *E. coli* citrate synthase is the unusual polypeptide chain folding found at the putative acetylcoenzyme A binding site. Key parts of this region, including His264 and a portion of polypeptide chain known from type I structures to form an adenine binding loop (residues 299−303), are shifted by as much as 10 Å from where they must be for substrate binding and catalysis to occur. Furthermore, the adjacent polypeptide chain composed of residues 267−297 is extremely mobile in our structure. Thus, acetylcoenzyme A binding to type II *E. coli* citrate synthase would require substantial structural shifts and a concerted refolding of the polypeptide chain to form an appropriate binding subsite. We propose that this essential rearrangement of the acetylcoenzyme A binding part of the active site is also a major feature of allostery in type II citrate synthases. Overall, this study suggests that the evolutionary development of hexameric association, the elaboration of a novel N-terminal domain, introduction of a NADH binding site, and the need to refold a key substrate binding site are all elements that have been developed to allow for the allosteric control of catalysis in the type II citrate synthases.

Citrate synthase (CS)[1] is an almost ubiquitous enzyme, found in all eukaryotes and in most, though not all, microorganisms. Its role as a catalyst of the entry point reaction for entry of two-carbon units into the citric acid cycle guarantees its importance for organisms with a complete cycle (Figure 1), but even where the complete cycle does not function, it is an essential step in the biosynthesis of amino acids related to glutamate. Some organisms have a single CS, while a few, such as baker's yeast and *Bacillus subtilis*, have as many as three. Most CS molecules are homodimers with a polypeptide chain between 380 and 440 residues in length (not including the signal peptide, when present). This kind of CS, which is designated type I, is found in eukaryotes, archaea, and Gram-positive bacteria. Detailed structures for vertebrate CS, obtained by X-ray crystallography, have been available for some time (*2*−*5*), and more recently, the type I structures of two CS from archaea (*6*, *7*) and one psychrophilic bacterium (*8*) were completed. These studies have permitted the detailed description of the active site region and allowed for proposals concerning the specific roles key residues have in substrate binding and catalysis (*3*, *9*). Support for these proposals has been obtained through subtle physical measurements (*10*−*13*) and by site-directed mutagenesis studies (*14*−*18*). The active site of each subunit is located between two independently folded domains, the

[†] This work was funded by an operating grant to H.W.D. and G.D.B. from the Canadian Institutes of Health Research.

[‡] Coordinates for the structure described in this work have been deposited in the Protein Data Bank (code 1K3P) (*1*).

* To whom correspondence should be addressed. Telephone: (604) 822-5216. Fax: (604) 822-5227. E-mail: brayer@laue.biochem.ubc.ca.

[§] University of British Columbia.

[||] These authors contributed equally to this work.

[⊥] University of Manitoba.

[1] Abbreviation: CS, citrate synthase. The amino acid numbering is according to the sequence of *E. coli* citrate synthase (see Table 2).

10.1021/bi010408o CCC: $20.00 © 2001 American Chemical Society
Published on Web 10/13/2001

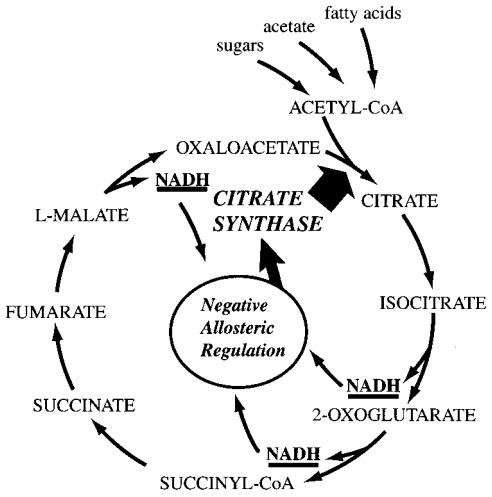

FIGURE 1: Schematic representation of the metabolic role of type II *E. coli* citrate synthase in the citric acid cycle, emphasizing the importance of NADH as a primary end product and the negative allosteric control it exerts over this enzyme which is the catalytic entry point to the cycle.

large and small domains. Comparison of vertebrate CS structures with and without substrates present has shown a clear-cut case of induced fit, such that substrate binding causes a partial closing of the two domains. The details of this transition between "open" and "closed" forms have been the subject of considerable analysis (*19*−*21*).

Much less structural information is available for the type II form of CS, and to this point, it has not been possible to determine the three-dimensional structure of an enzyme of this class. This second type of CS was first recognized by P. D. J. Weitzman in 1966 (*22*). Type II CS is a hexameric molecule, found only in Gram-negative bacteria (*23*−*25*). Its most striking functional property, which clearly distinguishes it from type I enzymes, is that type II CS is strongly and specifically inhibited by NADH via an allosteric mechanism (*22*, *24*, *26*, *27*; Figure 1). Furthermore, in the type II CS from *Escherichia coli*, the enzyme of this kind most extensively studied, the substrate saturation curve for one substrate, acetylcoenzyme A, is sigmoid in low-ionic strength buffers, but becomes hyperbolic in the presence of sufficient KCl and other salts. KCl also abolishes the NADH inhibition, thus acting like an allosteric activator (*22*, *27*, *28*). In other examples of type II CS, from *Acinetobacter anitratum* and *Pseudomonas aeruginosa*, the allosteric kinetics are somewhat different (*29*, *30*), but it has been suggested that all these enzymes employ the same allosteric mechanism. The differences observed may arise from differences in the value of the allosteric constant, *L*, the ratio between the T (inactive) and R (active) conformational states in the absence of ligands (*29*). The sequences of the polypeptide chains of type II CS are significantly identical (∼26%) to that of type I CS, so the allosteric properties of type II CS must represent an evolutionary refinement of a type I enzyme. Type II CS is therefore a potentially valuable model for understanding how regulatory properties evolve.

Notably, all the active site residues in type I CS have counterparts in the type II sequences. This point, first made from a comparison of the *E. coli* and pig heart CS sequences (*31*), is now abundantly confirmed by the many CS sequences that are now available. These sequences show that the original classification into "small" (dimeric) and "large" (hexameric) forms of CS could easily be further refined. For example, the type I CS enzymes of eukaryotes have important sequence differences from the type I of archaea, and from those of Gram-positive bacteria, though all are dimers as far as we know. Interestingly, some Gram-negative bacteria, in addition to their type II CS enzymes, have a gene for what seems to be a second, type I-like CS molecule (*32*, *33*). In *E. coli*, at least, this type I-like CS is actually a methylcitrate synthase, using propionylcoenzyme A as a substrate (*34*, *35*).

On the basis of the overall sequence similarity between type I and II CS enzymes, a model of the *E. coli* type II CS subunit had been previously constructed (*36*), and this has been used in functional studies of the active site of *E. coli* CS (*37*, *38*). However, this model can say nothing conclusive about the structural elements involved in hexameric type II enzyme formation, about the location of the NADH inhibitor binding site, or about the structural nature of the allosteric conformational change induced by NADH and how this influences active site residues. To address these critical mechanistic questions, we report here the first three-dimensional structure of a type II CS hexamer, that from *E. coli*.

EXPERIMENTAL PROCEDURES

E. coli citrate synthase was purified as previously described (*39*). Crystals of this enzyme were grown using the hanging drop vapor diffusion method (*40*) from 2 to 2.3 M ammonium sulfate, 2% (v/v) PEG 400, and 0.1 M Na Hepes at pH 6.0. These crystals reached dimensions of up to 0.8 mm × 0.7 mm × 0.3 mm. Diffraction data were collected from a single crystal on an R-Axis II area detector at 100 K, and then processed and scaled with the HKL (*41*) and CCP4 (*42*) program suites. The structure was determined by molecular replacement methods using the backbone structure of *Pyrococcus furiosus* citrate synthase (*7*) as the search model and a detwinning strategy based on the method of Yeates (*43*), as implemented in CNS (*44*). Rotation and translation searches were conducted using the program AMoRe (*45*). Subsequent rigid body refinement, simulated annealing procedures, and positional and thermal factor refinements with CNS allowed for the placement of most polypeptide chain residues. Upon application of a detwinning approach (perfect merohedral), in addition to the use of difference electron density maps over the entire polypeptide chain, as well as further positional and thermal parameter refinement, it was possible to position the remaining residues and complete the structural refinement (Table 1). Substantial positional disorder is observed for residues 1−9 and 267−297. Assessment of the refined structure of *E. coli* citrate synthase showed that it has excellent polypeptide chain geometry (Table 1). The coordinate error is estimated to be 0.24 Å (*46*).

RESULTS AND DISCUSSION

Hexamer Conformation. The overall hexamer structure of *E. coli* citrate synthase is best characterized as a torus with an ∼125 Å diameter and a thickness of ∼90 Å (Figure 2). It is constructed from three identical dimer units arranged around a 3-fold axis. Although the structure is not constrained by symmetry, within this structure there remains a strong

Table 1: Summary of Structure Determination Statistics

data collection parameters	
space group	*R*3
unit cell dimensions (Å)	
$a = b$	165.4
c	155.3
total no. of measurements	638153
no. of unique reflections	79578
mean $I/\sigma I$[a]	13.3 (6.7)
multiplicity[a]	3.2 (2.9)
merging *R*-factor (%)	9.3 (17.8)
maximum resolution (Å)	2.2
structure refinement values	
no. of reflections	78948
resolution range (Å)	10−2.2
completeness within range (%)[a]	98.4 (99.4)
no. of protein atoms	6732
no. of solvent atoms	876
average thermal factors (Å^2)	
protein atoms	42.7
solvent atoms	43.5
final R_{free} value (%)[b]	22.6
final crystallographic *R*-factor (%)	17.8
final structure stereochemistry	
rmsd for bonds (Å)	0.007
rmsd for angles (deg)	1.5

[a] Values in parentheses are for the highest-resolution shell (2.24−2.20 Å resolution). [b] The test set contained 4.7% of available data (3774 reflections).

2-fold relationship between the monomer units in each dimer (rms deviation for main chain atoms of 0.37 Å). However, as Figure 3 shows, there are regions of the polypeptide chain that take on significantly different conformations in each monomer, with the largest positional differences occurring in the vicinity of residues 141−164. This is markedly different from the dimeric type I enzymes where a defined 2-fold relationship is found between monomer units, and therefore, the conformations of both monomers in a dimer unit are identical.

The most prominent interactions holding the *E. coli* citrate synthase hexamer complex together are localized about a cationic porelike structure (~6 Å diameter) that runs along the central 3-fold axis relating dimer enzyme units (Figure 4). Each monomer, of each of the three citrate synthase dimers, has several segments of polypeptide chain that form interactions in this pore region. The majority of such contacts involve residues 111−119, 122−127, 178−190, and 204−207. A surprising feature of this central pore is the highly positively charged nature of its inner cylindrical surface, which is lined with arginine side chains (Figure 4). These side chains are contributed by Arg119, Arg125, and Arg126 (18 in total from the six monomer units), and are accompanied by an extensive network of associated hydrogen bonding and charged interactions. Most of the amino acids forming the cationic pore are strongly conserved among type II citrate synthases, but the equivalent region is very different in sequence in the type I enzymes (PIG, BAC, and PFU sequences in Table 2). This suggests the possibility that the arginine rich pore observed is closely linked to the regulatory aspects of the type II enzymes. Notably, *E. coli* citrate synthase is strongly activated by various salts in the concentration range of 10−200 mM (*28*), and this phenomenon may be connected with the observed pore structure. Thus, interactions between arginine side chains in the pore may force the enzyme into the T (inactive) conformational

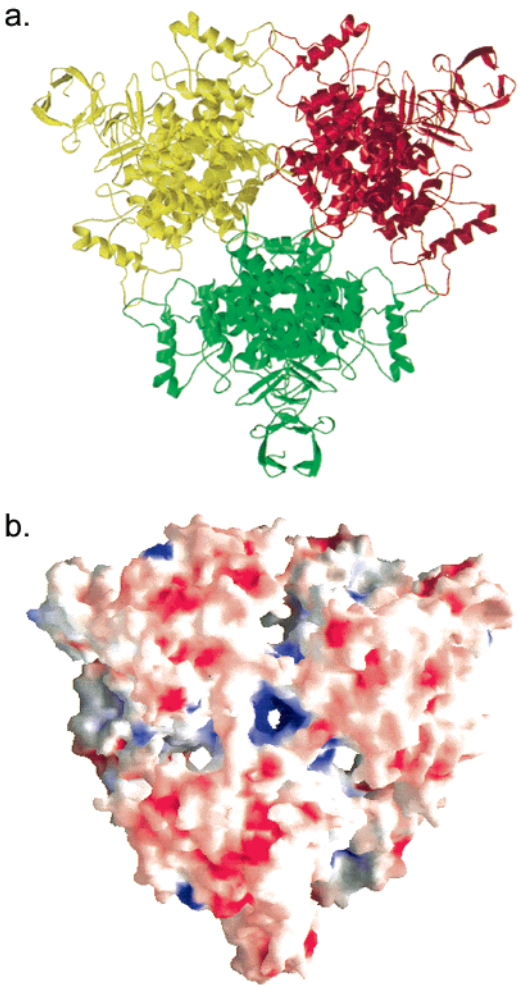

FIGURE 2: (a) Overall hexameric structure of type II *E. coli* citrate synthase shown using ribbons to represent the folds of the six polypeptide chains. The three equivalent dimer units related by a central 3-fold axis have been individually colored green, red, and yellow. (b) Electrostatic surface diagram of the full hexamer structure of type II *E. coli* CS demonstrating the highly positive nature of the central pore around which the majority of intermolecular hexamer interactions are concentrated. These figures were prepared with the assistance of Molscript (*65*), Raster3D (*66*), and GRASP (*67*).

state, and adding salt, by weakening these ionic interactions, might allow a relaxation into the R (active) state.

Subunit Fold and Homology with Other CS Structures. Apart from the 52 N-terminal amino acids (largely absent in nonmammalian type I CS; Table 2), the tertiary structure fold of the *E. coli* type II CS subunit is similar to that found for pig heart type I CS (*2*) and the type I CS from three different extremophiles (*6*−*8*). Superpositions of the main chain backbone of *E. coli* CS on the structures of pig heart and *P. furiosus* CS are shown in Figure 5. Another distinctive feature of *E. coli* type II CS is at the C-terminal end of the polypeptide chain, which folds in a manner similar to that of the type I *P. furiosus* CS and markedly different from that found in the mammalian type I pig heart CS. Other differences between these enzymes are primarily confined to loops and turns between helical segments, and to a disordered region in *E. coli* CS between residues 267 and 297.

Table 2 contains a sequence alignment that shows the locations of secondary structural elements in *E. coli* type II CS (labeled ECO) and the two other representative type I

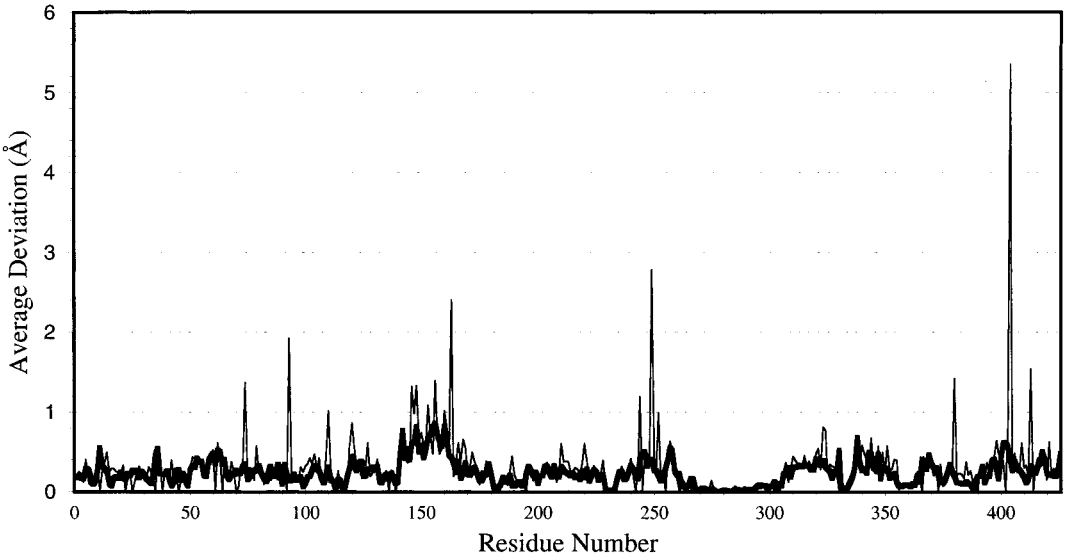

FIGURE 3: Average positional deviations between the main (thick lines) and side (thin lines) chain atoms of the two subunits in each dimer unit of *E. coli* citrate synthase. On the basis of the best fit of all main chain atoms, these subunits could be superimposed with a rmsd of 0.37 Å. A unique property of the type II allosterically controlled citrate synthases is the lack of a strict 2-fold relationship between subunits.

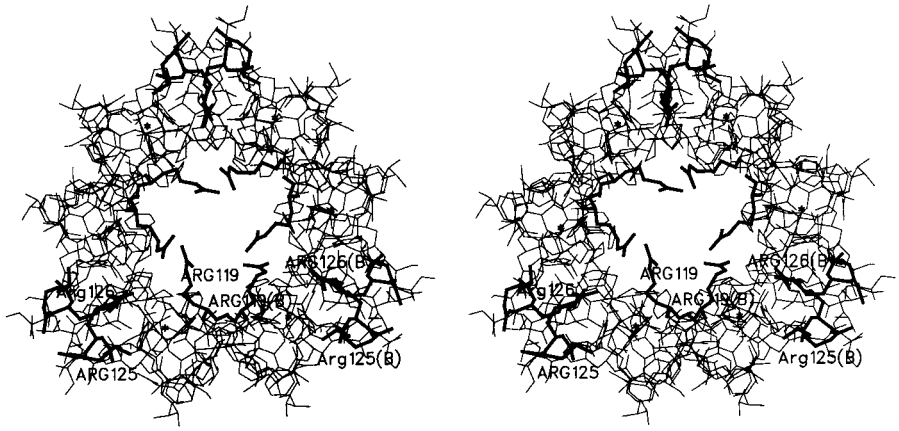

FIGURE 4: Detailed stereodrawing of the region about the central cationic pore found in the hexameric structure of type II *E. coli* citrate synthase. Six groups of arginines 119, 125, and 126 (shown as bold lines), from the six enzyme subunits in the hexameric complex (those of the A and B subunits of one dimer are labeled), form the primary interactions in this pore structure. These residues are in turn associated with an extensive network of hydrogen bond interactions with other amino acids. An asterisk has been placed at each of the locations of Cys206 in the type II *E. coli* citrate synthase hexamer to indicate its positioning and that of the related JK loop (Table 2).

structures shown in Figure 5 (labeled PIG and PFU in Table 2). Also included in Table 2 are the sequences of the other two well-characterized type II CS enzymes, those from *Acinetobacter calcoaceticus var. anitratus* (ACI) and *Ps. aeruginosa* (PSE), and the sequence of one representative Gram-positive Type I CS (*B. subtilis citA* gene product, labeled BSU). No tertiary structures are yet available for these three latter enzymes.

Figure 5 and Table 2 show that all the helices (labeled A−G and I−S) found in the *E. coli* CS subunit have counterparts in the other known CS structures, and the main differences between these structures occur at interhelical loops. For example, pig heart CS has nine more residues at the CD corner, and seven more residues between helices G and I (forming a helix H that is not present in the nonvertebrate examples) than the *E. coli* and *P. furiosus* enzymes do. The NO region is shorter in the *E. coli* enzyme by five residues, and in *P. furiosus* by seven residues, and the QR corner is longer by five residues in pig CS than in the others. On the other hand, the *E. coli* enzyme has two significant sequence insertions, relative to pig and *P. furiosus* CS, with seven more residues at the JK corner, and seven at the PQ corner. The JK insertion is especially significant because, as will be explained below, it is involved in dimer−dimer contacts, and it contains the reactive Cys206, a residue implicated in inhibition by the allosteric ligand, NADH.

The current structural results, showing that the subunits of *E. coli* type II CS are folded in a manner similar to that of pig type I CS, are consistent with the finding of earlier work which showed that limiting amounts of subtilisin leads to digestion of both enzymes at almost the same points in the two sequences. In the case of the *E. coli* enzyme, the most susceptible peptide bonds are those involving the amide nitrogens of Gly302, Gly304, and Val307. Somewhat slower to cleave were bonds in a second region, on either side of Lys356 (*47*). The first group of targets involves the predicted acetylcoenzyme A binding loop and the key active site residue His305, and the second pair of targets is in the QR helical corner. The most sensitive subtilisin target in pig CS is the Ala321−Val322 bond, in the OP nonhelical loop (*48*, *49*). As can be seen from Table 2, this position is exactly homologous to the Arg306−Val307 subtilisin target in *E.*

Table 2: Alignment of Primary and Secondary Structures for Selected Citrate Synthases[a]

```
             /-----------A----------\       /---B---\                |-|    |-|      /--C---\          /
PIG       ASSTNLKDILADLIPKEQARIKTFRQQHGNTVVGQITVDMMYGGMRGMKGLVYETSVLDPDEGIRFRGYSIPECQKMLPKAKGGEEPL  88
             |---| |---||-|        |--|          |--|       |-|    |---|     |---| |/|--C--\          /
ECO     ADTKAKLTLNGDTAVELDVLKGTLGQDVIDIRTLGSKGVFTFDPGFTSTASCESKITFIDGDEGILLHRGFPIDQLATDS---------N  81
        ^^^^^^^^^^^^^^
ACI   SEATGKKAVLHLDGKEIELPIYSGTLGPDVIDVKDVLASGHFTFDPGFMATASCESKITFIDGDKGILLHRGYPIDQLATQA---------D  83
PSE     ADKKAQLIIEGSAPVELPVLSGTMGPDVVDVRGLTATGHFTFDPGFMSTASCESKITFIDGDKGVLLHRGYPIEQLAEKS---------D  81
BAC                                              MVHYGLKGITCVETSISHIDGEKGRLIYRGHHAKDIALNH---------S  41
                                                                    |---|     |---| |/|--C---\          /
PFU                                        NTEKYLAKGLEDVYIDQTNICYIDGKEGKLYYRGYSVEELAELS---------T  45

    -----D----\    /-------E------\    /---F----\     /-------G------\/--H--\     /----------I-------
PIG PEGLFWLLVTGQIPTEEQVSWLSKEWAKRAALPSHVVTMLDNFPTNLHPMSQLSAAITALNSESNFARAYAEGIHRTKYWELIYEDCMDLIAK 181
    -----D---\      /------E------\  /----F----\     /-------G------\               /------I-------
ECO YLEVCYILLNGEKPTQEQYDEFKTTVTRHTMIHEQITRLFHAFRRDSHPMAVMCGITGALAAFY-------HDSLDVNNPRHREIAAFRLLSK 167
                                  * ***  *  *  ***
ACI YLEYCYLLLNGELPTAEQKVEFDAKVRAHTMVHDQVSRFFNGFRRDAHPMAIMVGVVGALSAFY-------HNNLDIEDINHREITAIRLIAK 169
PSE YLETCYLLLNGELPTAAQKEQFVGTIKNHTMVHEQLKTFFNGFRRDAHPMAVMCGVIGALSAFY-------HDSLDITNPKHREVSAHRLIAK 167
BAC FEEAAYLILFGKLPSTEELQVFKDKLAAERNLPEHIERLIQSLPNNMDDMSVVRTVVSALGENT-------------YTFHPKTEEAIRLIAI 121
    -----D----\    /-------E------\ /-----F----\     /-------G-------\               /------I-------
PFU FEEVVYLLWWGKLPSLSELENFKKELAKSRGLPKEVIEIMEALPKNTHPMGALRTIISYLGNID-------DSGDIPVTPEEVYRIGISVTAK 131

    ------I------\          /------J-----\           /-----K------\      /-----L---\    /----M----
PIG LPCVAAKIYRNLYREGSSIGAIDSKLDWSHNFTNMLGY--------TDAQFTELMRLYLTIHSDHEGGNVSAHTSHLVGSALSDPYLSFAAAM 266
    ------I------\            /----J-----\          /------K-------\     /-----L-----\  /----M----
ECO MPTMAAMCYKYSIGQP--FVYPRNDLSYAGNFLNMMFSTPCEP-YEVNPILERAMDRILILHADHEQ-NASTSTVRTAGSSGANPFACIAAGI 256
                ***     * *             ****
ACI IPTLAAWSYKYTVGQP--FIYPRNDLNYAENFLHMMFATPADRDYKVNPVPARAMDRIFTLHADHEQ-NASTSTVRLAGSTGANPYACISAGI 259
PSE MPTIAAMVYKYSKGEP--MMYPRNDLNYAENFLHMMFNTPCET-KPISPVPAKAMDRIFILHADHEQ-NASTSTVRLAGSSGANPFACIASSI 256
BAC TPSIIAYRKRWTRGEQ--AIAPSSQYGHVENYYYMLTG------EQPSEAKKKALETYMILATEHGM-NASTFSARVTLSTESDLVSAVTAAL 205
    ------I------\            /----J-----\          /------K--------\    /-----L------\ /----M----
PFU IPTIVANWYRIKNGLE--YVPPKEKLSHAANFLYMLHG------EEPPKEWEKAMDVALILYAEHEI-NASTLAVMTVGSTLSDYYSAILAGI 215

    -M-\    /-------N-------\     /-------O------\ .....        /------P------\          /--Q----
PIG MGLAGPLHGLANQEVLVWLTQLQKEVGKDVSDEKLRDYIWNTLNSGRVVPGYGHAVLRKTDPRYTCQRETALKHLP-------HDPMFKLVAQ 352
    --M-\                                                      /------P-------\         /--Q----
ECO ASLWGPAHGGANEAALKMLEEISSVKH-----IPEVVRRAKDKNDSFRLMGFGHRVYKNYDPRATVMRETCHEVLKELGTKDDLLEVAMELEN 344
         ^^*^^^^^^^^^^^^^^^^^     ^^^^^^^^^^^^^^^^                              ^^^^^^^^^^^^
ACI SALWGPAHGGANEAVLKMLDEIGSVEN-------VAEFMEKVKRKEVKLMGFGHRVYKNFDPRAKVMKQTCDEVLEALGINDPQLALAMELER 345
PFU AALWGPAHGGANEAVLRMLDEIGDVSN-----IDKFVEKAKDKNDPFKLMGFGHRVYKNFDPRAKVMKQTCDEVLQELGINDPQLELAMKLEE 344
BAC GTMKGPLHGGAPSAVTKMLEDIGEKEH-------AEAYLKEKLEKGERLMGFGHRVYKTKDPRAEALRQKAEEVAGNDRDLDLALHVEAEAIR 291
    -M-\     /-----N------\          /----O------\              /------P------\       /----Q-----
PFU GALKGPIHGGAVEEAIKQFMEIGSPE-------KVEEWFFKALQQKRKIMGAGHRVYKTYDPRARIFKKYASKLG-------DKKLFEIAERL 294

    ----Q-------\         /-----R-----\  /------------S------------\         /-----T----\
PIG LYKIVPNVLLEQGKAKNPWPNVDAHSGVLLQYYGMTEMNYYTVLFGVSRALGVLAQLIWSRALGFPLERPKSMSTDGLIKLVDSK          437
    ----Q------\         /------R-----\  /------------S-------------\
ECO IALNDPYFIEKK-----LYPNVDFYSGIILKAMGI-PSSMFTVIFAMARTVGWIAHWSEMHSDGMKIARPRQLYTGYEKRDFKSDIKR      426
                                                                                          ^^^
ACI IALNDPYFVERK-----LYPNVDFYSGIILKAIGI-PTEMFTVIFALARTVGWISHWLEMHSKPYKIGRPRQLYTGEVQR----DIKR      423
PSE IARHDPYFVERN-----LYPNVDFYSGIILKAIGI-PTSMFTVTFALARTVGWISHWQEMLSGPYKIGRPRQLYTGHTQRDFTALKDRG    427
BAC LLEI--YKPGRK-----LYTNVEFYAAAVMRAIDF-DDELFTPTFSASRMVGWCAHVLEQAENNMIF-RPSAQYTGAIPEEVLS          366
    ----Q--\               /---R----\     /----------S------------\
PFU ERLVEEYLSKKG-----ISINVDYWSGLVFYGMKI-PIELYTTIFAMGRIAGWTAHLAEYVSHN-RIIRPRLQYVGEIGKKYLPIELRR     376
```

[a] A representative nonallosteric mammalian type I CS: PIG, pig heart (*48*, *49*). Allosteric type II enzymes from Gram-negative bacteria: ECO, *E. coli* (*31*, *68*); ACI, *A. anitratum* (*69*); and PSE, *Ps. aeruginosa* (*30*). Other nonallosteric type I enzymes from a Gram-positive bacterium and an archaeon: BAC, *B. subtilis* citrate synthase I [*citA* product (*70*); Swiss Prot entry P39119]; and PFU, *P. furiosus* CS (*7*). At the end of each sequence line, the number of the last residue is given. The color code is as follows: red, active site residues; blue, additional residues conserved in all CS sequences; and green, other residues also conserved in allosteric CS sequences. A series of five black dots mark residues 299−303 (*E. coli* CS numbering), the main chain atoms of which bind the adenine ring of coenzyme A in pig CS. Regions of α-helix are indicated above the respective sequences for the three enzymes for which tertiary structures are known, as a series of dashes, with the first and last residues labeled with oblique strokes. The letter assigned to each helix follows the conventions of Remington et al. (*2*). *β*-Strands are similarly indicated but, in this case, with the first and last residues labeled with vertical bars. Highly mobile regions of *E. coli* CS are underlined with carat marks. Residues involved in dimer−dimer contacts in the *E. coli* enzyme are denoted with asterisks below the line. In the *E. coli* enzyme, the three domains of each subunit consist of residues 1−52 (N-terminal *β*-sheet domain), 53−262 and 375−426 (large domain), and 263−374 (small domain).

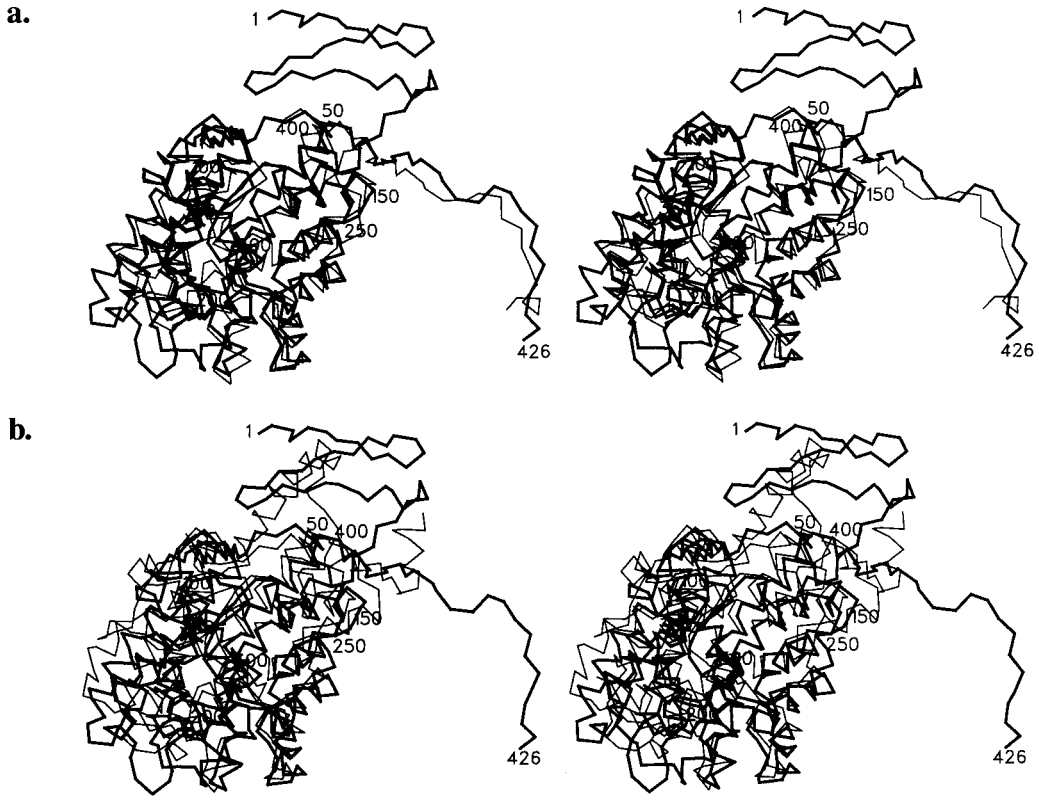

FIGURE 5: Stereodrawings showing the superposition of the polypeptide chain backbone of type II *E. coli* CS (thick lines) on the backbone structures of the type I CS enzymes (thin lines) from (a) *P. furiosus* and (b) pig. Every 50th amino acid in the sequence of *E. coli* CS (Table 2) has been labeled, along with the N- and C-terminal ends of this enzyme.

coli CS. In a more extensive study of the effects of other proteases on pig CS, Lill et al. (*50*) showed that, like subtilisin, most proteases prefer bonds in the OP loop. LysC protease, on the other hand, attacks the Lys366−Lys367 bond selectively, a site exactly homologous to the Lys355−Lys356 bond, the second subtilisin target in *E. coli* CS (Figure 5). In the pig enzyme, these target bonds are in surface loops. Our *E. coli* CS structure now shows that, despite the potential congestion caused by hexamer formation, the subtilisin targets remain relatively accessible on the enzyme surface.

Novel E. coli CS N-Terminal Domain. The first ~52 amino acids of each of the *E. coli* CS subunits associated as a dimer jointly combine to form a single compact and independently folded domain rich in β-sheet structure. Contacts between this N-terminal domain and the rest of the protein are very limited. A schematic diagram showing in detail the folding of this domain is shown in Figure 6A, and a ribbon drawing showing its overall tertiary structure is shown in Figure 6B. Notably, the three β-sheets present retain the approximate 2-fold symmetry that exists between the two subunits of the dimer structure (Figure 3). Also of interest is the fact that the polypeptide chains of both monomers participate in all three β-sheets of this N-terminal domain. If the two subunits are designated A and B, the three β-sheets can be described as follows: (1) antiparallel, consisting of residues 15A−17A, 11A−7A, 7B−11B, and 17B−15B; (2) mixed parallel/antiparallel, consisting of residues 18B−22B, 31B−26B, 43A−40A, and 50B−48B; and (3) mixed parallel/antiparallel, consisting of residues 18A−22A, 31A−26A, 43B−40B, and 50A−48A.

Earlier, automated sequence alignment procedures carried out as part of the Pfam project (*51*) had identified the N-terminal sequence of *E. coli* CS as a potential domain structure (it is designated Pfam B_3152), and this domain is likely present in all CS sequences from Gram-negative bacteria. The function of this novel, well-conserved domain is not known, but its location in the hexameric *E. coli* CS structure is remote from both the dimer−dimer contact regions and catalytic active sites (Figures 2 and 5); therefore, it probably does not play a role in hexamer formation.

There is no sequence corresponding to this N-terminal domain in type I CS sequences from Gram-positive bacteria or archaea (see Table 2 sequences BAC and PFU, and Figure 5), which therefore are ~40−50 residues shorter. Furthermore, the N-terminal domain found in *E. coli* CS is absent from a second type I citrate synthase-like protein of *E. coli*, which has been shown to be a methylcitrate synthase (*34*, *35*), and also from a second type I CS of *Ps. aeruginosa* (*33*, *52*).[2] Eukaryotic type I CS sequences do have an N-terminal extension, but in the known vertebrate structures, this is folded completely differently, largely as a long helix A that lies on the outer surface of the subunit (Figure 5). Other eukaryotic type I CS molecules, whose sequences are highly homologous to vertebrate CS in the N-terminal region, are likely folded in the same way.

Subunit−Subunit Contacts. With the exception of the additional contacts within the N-terminal domain, the subunits forming each dimer of *E. coli* type II CS interact in a manner similar to those of other known type I CS

[2] What is regarded as the principal *Ps. aeruginosa* CS gene, designated *gltA*, is gene PA1580 in section 147 of the *Ps. aeruginosa* genome project; the second (shorter) CS-like gene, designated *prpC* or PA0795, is in genome section 75 (*52*).

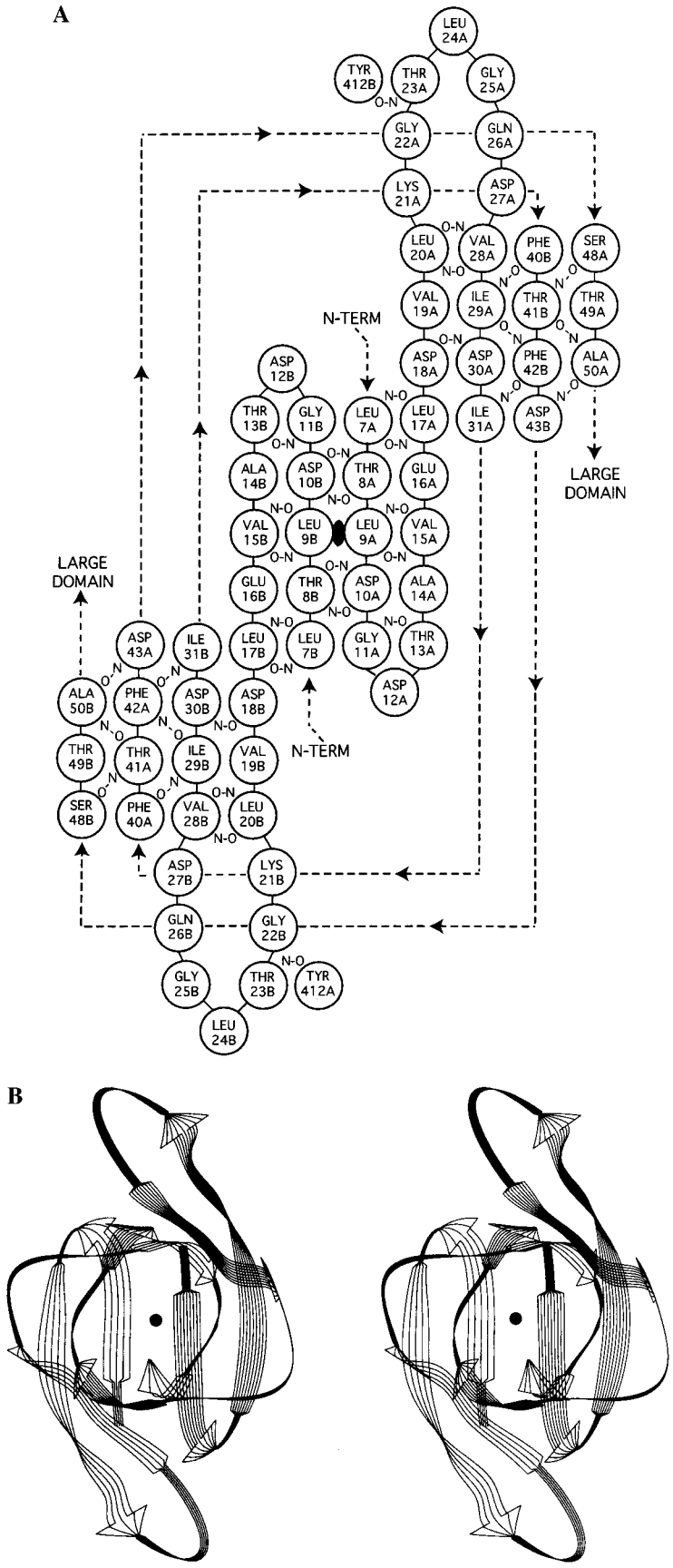

FIGURE 6: Analysis of the unique N-terminal domain of type II *E. coli* CS. (A) A schematic drawing of the polypeptide chain folding pattern and main chain hydrogen bonding formed in the three β-sheets present. This domain is made up of the two N-terminal ends (residues 1−52, chains A and B) of the two subunits that make up a dimer unit of type II *E. coli* CS. (b) Closeup stereodrawing emphasizing the compact nature of this domain and the packing arrangement of its β-sheet components. The two polypeptide chains present are distinguished by narrower and wider striped arrows, and the large black dot denotes the approximate 2-fold axis that runs through this domain. Sequence comparisons suggest that this domain structure is only found in the type II CS (Table 2).

structures. This contact surface involves the large domain, where each of helices F, G, L, and M interacts with its counterpart in the other subunit. In terms of individual subunit active sites, these are formed predominantly by residues within a particular subunit. However, one key residue which assists in OAA binding, Arg407, is contributed from the other subunit of the dimer. Overall, the interaction between subunits is strong, and there is no evidence that any CS can dissociate into its individual subunits in the absence of denaturation.

Dimer−Dimer Interactions. These are a requirement for hexamer formation and a novel feature of the *E. coli* type II CS. The dimer−dimer contact surfaces that are formed involve residues from corner EF, helix F, corner FG, and corner IJ (Table 2). All of these residues are localized in one relatively small region of the large domain. In addition, residue Glu276 becomes slightly more buried upon hexamerization, but this is in a region of considerable motional disorder; therefore, the evidence for this interaction is not strong. Most dimer−dimer interactions are van der Waals contacts, but there is one good hydrogen bond between the carbonyl oxygen of Cys206 in one dimer and the main chain amide nitrogen of Asn189 of the other. The residues that show the greatest loss of the level of solvent exposure upon hexamerization (<15% of their surface remains exposed) are Arg126, Gly181, Pro187, and Cys206.

Cys206 is an especially interesting residue since its chemical modification abolishes inhibition by the allosteric effector, NADH (*26*, *53*), and this correlates with a loss of hexamer formation (*54*). Moreover, Cys206 is located in the midst of a sequence insertion involving loop JK (Table 2), which is characteristic of type II CS. The exact size of this insertion is somewhat variable. In most type II CS sequences reported to date, including the known allosteric enzymes from *E. coli* and *P. aeruginosa*, there are seven extra residues, whereas in the allosteric enzyme from *A. anitratus*, there are eight. For the type II CS from *Helicobacter pylori*, loop JK is 15 residues long (alignment not shown).[3] This enzyme is not inhibited by NADH (*55*), and its subunit structure has not been established.

E. coli CS Active Site. Although our structure contains no active site ligands, the active site is easily located because of its high degree of homology to the active site of type I CS. The amino acids concerned with substrate binding by type I CS, and presumably catalysis, have been defined by extensive structural studies, employing a number of complexes with citrate, OAA, and various coenzyme A analogues (*3*, *4*, *9*). Equivalent residues were recognized in structural studies of extremophile CS (*6*−*8*), and alignments of the many other CS sequences that are now available show that all the key residues are fully conserved throughout the series. In the representative sequences shown in Table 2, the key active site residues appear in red. In the *E. coli* enzyme, the roles of most of these residues have been verified by site-directed mutagenesis (*37*, *38*, *57*, *58*).

As with type I CS from pig, the kinetic mechanism of type II CS from *E. coli* appears to be ordered, with OAA binding first (*12*, *38*). The putative OAA binding residues in the active site of our type II CS structure are in positions

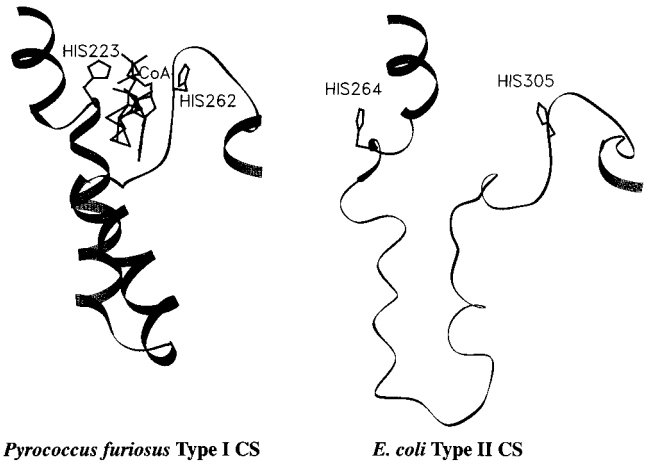

FIGURE 7: Polypeptide chain folding in the vicinity of the coenzyme A binding site in type I *P. furiosus* citrate synthase (with bound coenzyme A drawn) and type II *E. coli* citrate synthase (unliganded). Substantial refolding of residues 267−297 would be required in type II *E. coli* citrate synthase to form this binding site, and this process is likely linked to the unique allosteric properties of this enzyme.

essentially equivalent to those of their counterparts in a type I CS−OAA complex and other enzyme−substrate complexes (*4*, *6*−*8*, *59*). The only significant adjustment expected to facilitate OAA binding in *E. coli* CS would be movement of the side chain of Arg407. Arg407 is expected to form two hydrogen bonds to the γ-carboxyl of the substrate. This slightly more open arrangement of the OAA part of the active site, in *E. coli* type II CS, probably explains why the next higher homologous α-keto acid, 2-oxoglutarate, is an inhibitor of this enzyme, competitive with OAA (*37*, *60*). On the basis of the current structure determination, we do not expect that OAA binding to *E. coli* CS will induce a dramatic conformational change, and this is consistent with kinetic and binding studies that indicate that OAA is not an allosteric ligand.

In contrast, some parts of the active site of type II CS that are concerned with binding of the second substrate, acetylcoenzyme A, are in radically different positions (Figure 7 and Table 3). The most striking difference is the location of the side chain of His264, believed to act as the proton donor for enolization of the acetyl group, a key step in catalysis (*4*, *9*). In our structure, the imidazole group of His264 is ∼11 Å away from its required location for catalysis. In addition, residues 299−303, a strongly conserved sequence that wraps around the adenine ring of coenzyme A and forms the binding site for that part of the substrate (labeled with dots in Table 2), are removed from the required position to participate in coenzyme A binding. Interestingly, the main chain atoms comprising residues 267−297 have unusually high thermal factors (as much as 9 times the average; see Figure 8), indicating considerable mobility in this region, at least when ligands are absent. A key part of the allosteric mechanism of type II citrate synthases, as inferred from kinetic and binding data, is a conformational change induced by binding of acetylcoenzyme A (*25*, *27*, *61*). Our structure now suggests that this conformational change involves an extensive refolding of the polypeptide chain adjacent to, and within, the acetylcoenzyme A binding site.

[3] The CS gene of *H. pylori*, designated *gltA* or HP0026, is in section 3 of the *Helicobacter pylori* genome project (see ref *56*).

Table 3: Pairwise Average Positional Deviations (Å) of Active Site Residues in *E. coli* Type II CS from Those of Type I CS Structures

E. coli type II CS active site residues [in both independent subunits (A and B)]		comparison type I CS structures[a] (side chain/main chain)	
		pig	*P. furiosus*
His229	A	1.76/1.22	2.73/1.87
	B	2.08/1.14	2.87/1.59
His264	A	11.78/10.43	11.66/10.29
	B	10.98/10.76	10.94/10.01
His305	A	6.42/6.33	4.39/5.26
	B	5.98/6.07	4.86/4.87
Arg314	A	2.90/1.73	2.30/2.20
	B	2.97/1.63	2.09/2.18
Asp362	A	3.28/2.79	3.71/4.32
	B	2.98/2.67	3.87/4.54
Phe383	A	4.53/0.69	5.38/0.92
	B	5.12/0.78	5.64/0.78
Arg387	A	1.38/0.72	1.79/0.59
	B	1.45/0.56	1.97/0.63
Arg407	A	5.97/1.79	6.01/1.09
	B	6.39/1.58	5.87/0.94

[a] Based on the superposition of all main chain atoms between structures (see Figure 5). In the type I structures, both subunits of the dimer formed have the same conformation.

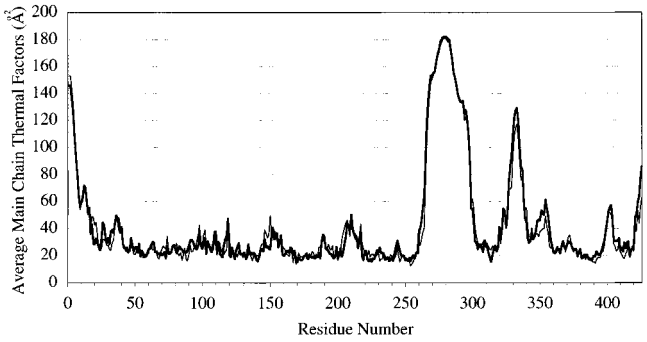

FIGURE 8: Plot of the average main chain thermal factors for the A (thick lines) and B (thin lines) subunits of a type II *E. coli* CS dimer. Unlike the strict 2-fold symmetry found for dimer subunits in type I enzymes, in the type II *E. coli* CS dimer this 2-fold relationship is only approximate, and therefore, the folding and thermal factor profiles of the two subunits vary. Note the large region of polypeptide chain (residues 267−297) having high positional mobility in both subunits of the *E. coli* enzyme. As illustrated in Figure 7, substantial refolding of this portion of the polypeptide chain would be required for catalytic activity to occur. This refolding event may be linked to the allosteric properties of the acetyl-CoA binding site of the type II CS.

Allostery and Type II E. coli CS. The current understanding of type II CS allostery, based primarily on kinetic and binding studies on the *E. coli* enzyme (*25*, *27*), is that there are two conformational states, one active (R state, binding acetylcoenzyme A, favored by high concentrations of KCl and other salts) and the other inactive (T state, binding NADH). Since OAA does not show sigmoid kinetics, and does not inhibit NADH binding by wild-type *E. coli* CS, its part of the active site was not expected to change on going from the T to R state. The fact that the acetylcoenzyme A binding site is in need of considerable refolding would suggest that our structure is in the T state, or at least that the T to R conformational change involves this part of the enzyme. Clearly, it would be desirable to obtain structures with NADH and acetylcoenzyme A bound, and we are now pursuing this with the wild-type enzyme and variants whose kinetics suggest substantial shifts in the T ↔ R equilibrium.

Several "allosteric" variant CS proteins have been reported, and their locations must indicate regions whose conformations are different in the T and R states. One, the Arg319Leu variant, has kinetic properties indicating a considerable shift toward the R state (*62*). The side chain of Arg319, in our structure, is in the small domain in the middle of helix P, where it makes two good H-bonds with the carboxyl group of Glu343 in helix Q. This interaction was predicted from modeling studies, but the Glu343Ala mutant shows little change in allosteric properties, a finding that remains unexplained. The Phe383Ala mutant, involving a highly conserved residue in the acetylcoenzyme A part of the active site, is strongly shifted to the T state (*38*). Only two residues within the highly mobile region composed of residues 267−297 have so far been mutated. One, Met274Ala, has kinetics similar to those of the wild-type enzyme, but the other, Leu275Ala, has unexpectedly low activity, only partly explained by its lower affinity for acetylcoenzyme A (*39*). The functional changes observed for this variant protein are consistent with a key role for the mobile loop of residues 267−297 in an allosteric conformational change.

Although the dimer−dimer contact region is relatively small, and *E. coli* CS hexamers do dissociate partially to dimers at low ionic strengths, it is important to emphasize that both the R and T states must be hexamers. NADH, the defining ligand for the T state, binds selectively to the hexamer conformation, and induces a shift in the dimer−hexamer equilibrium toward hexamer (*54*). There is little information available yet about the effect of substrates on this equilibrium, but KCl, the allosteric activator, also induces a shift toward hexamer (*63*).

E. coli CS reacts with a number of sulfhydryl group reagents, such as DTNB, and modification is accompanied by a loss of NADH inhibition (*26*). This reaction stops when one thionitrobenzoate anion has been released per CS subunit, and the reactive group has been identified as Cys206 (*53*, *64*). The reactivity is slightly greater in the presence of the activator, KCl, and it is considerably lower in the presence of NADH. Notably, a number of adenylates, such as AMP, ADP, and ADPribose, which are partial structural analogues of NADH, but not inhibitors of CS, do inhibit reaction of Cys206 with DTNB (*64*). These results indicate that Cys206 is in or very near the adenylate part of the NADH binding site, but that the dihydronicotinamide-binding part of this site must be occupied for enzyme inhibition to occur. Our current structure locates Cys206 in the all-important JK loop, at the heart of the dimer−dimer interface (Figure 4). This suggests the NADH site must be in this region, and the fact that NADH binds selectively to hexamer indicates that its site is not fully assembled until dimers associate into a hexameric conformation (*54*). The fact that KCl, the allosteric activator, slightly increases Cys206 reactivity (*64*) also indicates that the T → R conversion involves some rearrangement at this part of the site.

One puzzling aspect of Cys206 reactivity that is not resolved by the *E. coli* CS structure is the fact that, once modified by DTNB to form what is presumably the mixed Cys206−TNB disulfide, the protein slowly loses an additional 1 equiv of TNB anion. This process occurs over a period of many hours, and it is slower yet in the presence of 0.1 M KCl. It has been speculated that a second sulfhydryl

group is located close enough to Cys206 to displace the TNB, forming an intraprotein disulfide bond (*64*). Good analytical data were obtained to support this proposal, but from examination of the current structure, it is clear that there are no other sulfhydryl groups close to Cys206. Further experiments, using the structural information that is now available, will be needed to resolve this matter.

To conclude, the structure reported here has been invaluable in defining those parts of the type II CS dimer that interact with each other to produce the characteristic hexameric form, and in allowing us to understand the sequence differences between type I and II enzymes in a structural context. The observation of a central cationic pore, and the discovery of the N-terminal domain, are notable surprises. The finding that large portions of the acetylcoenzyme A binding site are in a considerably different conformation from what would be needed for substrate binding, and that the polypeptide chain involved is highly mobile, provides a plausible explanation for at least one aspect of type II CS allosteric properties. Interestingly, this rearrangement of the acetylcoenzyme A site is reminiscent of the open state of vertebrate CS, wherein the same parts of the active site are affected, and we might provisionally describe our structure as a "super open" state. That is, the open → closed transition in type I CS, a useful way of enclosing substrates so that they can react in a protected environment, seems to have been recruited and extended in type II enzymes to allow development of allosteric control over citrate synthase catalysis.

ACKNOWLEDGMENT

We thank Gunnar Olovsson and Gary Sidhu for insightful discussions and Michael Danson and Garry Taylor for making the coordinates for *P. furiosus* citrate synthase available to us. Figures 4 and 5 were prepared using SETOR (*71*), and Figures 6b and 7 were constructed with Molscript (*65*).

REFERENCES

1. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
2. Remington, S. J., Wiegand, G., and Huber, R. (1982) *J. Mol. Biol. 158*, 111−152.
3. Wiegand, G., and Remington, S. J. (1986) *Annu. Rev. Biophys. Biophys. Chem. 15*, 97−117.
4. Karpusas, M., Branchaud, B., and Remington, S. J. (1990) *Biochemistry 29*, 2213−2219.
5. Liao, D. I., Karpusas, M., and Remington, S. J. (1991) *Biochemistry 30*, 6031−6036.
6. Russell, R. J., Hough, D. W., Danson, M. J., and Taylor, G. L. (1994) *Structure 2*, 1157−1167.
7. Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J., and Taylor, G. L. (1997) *Biochemistry 36*, 9983−9994.
8. Russell R. J., Gerike, U., Danson, M. J., Hough, D. W., and Taylor, G. L. (1998) *Structure 6*, 351−361.
9. Remington, S. J. (1992) *Curr. Top. Cell. Regul. 33*, 209−229.
10. Kurz, L. C., Ackerman, J. J., and Drysdale, G. R. (1985) *Biochemistry 24*, 452−457.
11. Kurz, L. C., and Drysdale, G. R. (1987) *Biochemistry 26*, 2623−2627.
12. Kurz, L. C., Shah, S., Crane, B. R., Donald, L. J., Duckworth, H. W., and Drysdale, G. R. (1992) *Biochemistry 31*, 7899−7907.
13. Gu, Z., Drueckhammer, D. G., Kurz, L., Liu, K., Martin, D. P., and McDermott, A. (1999) *Biochemistry 38*, 8022−8031.
14. Zhi, W., Srere, P. A., and Evans, C. T. (1991) *Biochemistry 30*, 9281−9286.
15. Kurz, L. C., Drysdale, G. R., Riley, M. C., Evans, C. T., and Srere, P. A. (1992) *Biochemistry 31*, 7908−7914.
16. Evans, C. T., Kurz, L. C., Remington, S. J., and Srere, P. A. (1996) *Biochemistry 35*, 10661−10672.
17. Kurz, L. C., Roble, J. H., Nakra, T., Drysdale, G. R., Buzan, J. M., Schwartz, B., and Drueckhammer, D. G. (1997) *Biochemistry 36*, 3981−3990.
18. Kurz, L. C., Nakra, T., Stein, R., Plungkhen, W., Riley, M., Hsu, F., and Drysdale, G. R. (1998) *Biochemistry 37*, 9724−9737.
19. Lesk, A. M., and Chothia, C. (1984) *J. Mol. Biol. 174*, 175−191.
20. Hayward, S., and Berendsen, H. J. (1998) *Proteins 30*, 144−154.
21. Hinsen, K., Thomas, A., and Field, M. J. (1999) *Proteins 34*, 369−382.
22. Weitzman, P. D. J. (1966) *Biochim. Biophys. Acta 128*, 213−215.
23. Weitzman, P. D. J., and Jones, D. (1968) *Nature 219*, 270−272.
24. Weitzman, P. D. J., and Dunmore, P. (1969) *Biochim. Biophys. Acta 171*, 198−200.
25. Weitzman, P. D. J., and Danson, M. J. (1976) *Curr. Top. Cell. Regul. 10*, 161−204.
26. Danson, M. J., and Weitzman, P. D. J. (1973) *Biochem. J. 135*, 513−524.
27. Duckworth, H. W., and Tong, E. K. (1976) *Biochemistry 15*, 108−114.
28. Faloona, G. R., and Srere, P. A. (1969) *Biochemistry 8*, 4497−4503.
29. Morse, D., and Duckworth, H. W. (1980) *Can. J. Biochem. 58*, 696−706.
30. Donald, L. J., Molgat, G. F., and Duckworth, H. W. (1989) *J. Bacteriol. 171*, 5542−5550.
31. Ner, S. S., Bhayana, V., Bell, A. W., Giles, I. G., Duckworth, H. W., and Bloxham, D. P. (1983) *Biochemistry 22*, 5243−5249.
32. Patton, A. J., Hough, D. W., Towner, P., and Danson, M. J. (1993) *Eur. J. Biochem. 214*, 75−81.
33. Mitchell, C. G., and Anderson, S. C. (1996) *Biochem. Soc. Trans. 24*, 46S.
34. Gerike, U., Hough, D. W., Russell, N. J., Dyall-Smith, M. L., and Danson, M. J. (1998) *Microbiology 144*, 929−935.
35. Textor, S., Wendisch, V. F., De Graaf, A. A., Muller, U., Linder, M. I., Linder, D., and Buckel, W. (1997) *Arch. Microbiol. 168*, 428−436.
36. Duckworth, H. W., Anderson, D. H., Bell, A. W., Donald, L. J., Chu, A. L., and Brayer, G. D. (1987) *Biochem. Soc. Symp. 54*, 83−92.
37. Anderson, D. H., and Duckworth, H. W. (1988) *J. Biol. Chem. 263*, 2163−2169.
38. Pereira, D. S., Donald, L. J., Hosfield, D. J., and Duckworth, H. W. (1994) *J. Biol. Chem. 269*, 412−417.
39. Ayed, A., and Duckworth, H. W. (1999) *Protein Sci 8*, 1116−1126.
40. McPherson, A. (1982) *Preparation and Analysis of Protein Crystals*, John Wiley & Sons, New York.
41. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
42. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
43. Yeates, T. O. (1997) *Methods Enzymol. 267*, 344−358.
44. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
45. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
46. Luzzati, P. V. (1952) *Acta Crystallogr. 5*, 802−810.

47. Bell, A. W., Bhayana, V., and Duckworth, H. W. (1983) *Biochemistry 22*, 3400−3405.
48. Bloxham, D. P., Ericsson, L. H., Titani, K., Walsh, K. A., and Neurath, H. (1980) *Biochemistry 19*, 3979−3985.
49. Bloxham, D. P., Parmelee, D. C., Kumar, S., Wade, R. D., Ericsson, L. H., Neurath, H., Walsh, K. A., and Titani, K. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 5381−5385.
50. Lill, U., Schreil, A., Henschen, A., and Eggerer, H. (1984) *Eur. J. Biochem. 143*, 205−212.
51. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. (2000) *Nucleic Acids Res. 28*, 263−266.
52. Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K., Wu, Z., and Paulsen, I. T. (2000) *Nature 406*, 959−964.
53. Donald, L. J., Crane, B. R., Anderson, D. H., and Duckworth, H. W. (1991) *J. Biol. Chem. 266*, 20709−20713.
54. Ayed, A., Krutchinsky, A., Ens, W., Standing, K. G., and Duckworth, H. W. (1998) *Rapid Commun. Mass Spectrom. 12*, 339−344.
55. Pitson, S. M., Mendz, G. L., Srinivasan, S., and Hazell, S. L. (1999) *Eur. J. Biochem. 260*, 258−267.
56. Tomb, J.-F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Karp, P. D., Smith, H. O., Fraser, C. M., and Venter, J. C. (1997) *Nature 388*, 539−547.
57. Handford, P. A., Ner, S. S., Bloxham, D. P., and Wilton, D. C. (1988) *Biochim. Biophys. Acta 953*, 232−240.
58. Man, W. J., Li, Y., O'Connor, C. D., and Wilton, D. C. (1994) *Biochem. J. 300*, 765−770.
59. Wiegand, G., Remington, S., Deisenhofer, J., and Huber, R. (1984) *J. Mol. Biol. 174*, 205−219.
60. Wright, J. A., Maeba, P., and Sanwal, B. D. (1967) *Biochem. Biophys. Res. Commun. 29*, 34−38.
61. Talgoy, M. M., and Duckworth, H. W. (1979) *Can. J. Biochem. 57*, 385−395.
62. Anderson, D. H., Donald, L. J., Jacob, M. V., and Duckworth, H. W. (1991) *Biochem. Cell Biol. 69*, 232−238.
63. Tong, E. K., and Duckworth, H. W. (1975) *Biochemistry 14*, 235−241.
64. Talgoy, M. M., Bell, A. W., and Duckworth, H. W. (1979) *Can. J. Biochem. 57*, 822−833.
65. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
66. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol. 277*, 505−524.
67. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins: Struct., Funct., Genet. 11*, 281−296.
68. Bhayana, V., and Duckworth, H. W. (1984) *Biochemistry 23*, 2900−2905.
69. Donald, L. J., and Duckworth, H. W. (1987) *Biochem. Cell Biol. 65*, 930−938.
70. Jin, S., and Sonenshein, A. L. (1994) *J. Bacteriol. 176*, 4669−4679.
71. Evans, S. V. (1993) *J. Mol. Graphics 11*, 134−138.

BI010408O