Impact of Phenolic Substrate and Growth Temperature on the *Arthrobacter chlorophenolicus* Proteome

Maria Unell[1], Paul E. Abraham[2], Manesh Shah[3], Bing Zhang[4], Christian Rückert[5], Nathan C. VerBerkmoes[2] & Janet K. Jansson[1,6*]

[1] Department of Microbiology, Swedish University of Agricultural Sciences, Box 7025, 750 07 Uppsala, Sweden.

[2] Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.

[3] Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.

[4] Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37232, USA.

[5]Institute for Genome Research and Systems Biology, Center for Biotechnology, Bielefeld University D-33594, Bielefeld, Germany.

[6] Lawrence Berkeley National Laboratory, Earth Sciences Division, 1 Cyclotron Road, MS 70A, Berkeley, CA, 94720 USA.

[*]Corresponding author:

Janet K. Jansson:

Phone (510) 486-7487
Fax (510 486-7152
e-mail: jrjansson@lbl.gov

Keywords: *Arthrobacter chlorophenolicus*, 4-chlorophenol, 4-nitrophenol, phenol, shotgun proteomics, 2D-LC-ES-MS/MS, hydroxyquinol dioxygenase

Abstract

We compared the *Arthrobacter chlorophenolicus* proteome during growth on 4-chlorophenol, 4-nitrophenol or phenol at 5°C and 28°C; both for the wild type and a mutant strain with mass spectrometry based proteomics. A label free workflow employing spectral counting identified 3749 proteins across all growth conditions, representing over 70% of the predicted genome and 739 of these proteins form the core proteome. Statistically significant differences were found in the proteomes of cells grown under different conditions including differentiation of hundreds of unknown proteins. The 4-chlorophenol-degradation pathway was confirmed, but not that for phenol.

Introduction

Arthrobacter chlorophenolicus is a previously described species that is capable of degrading several *para*-substituted phenolic compounds, such as 4-chlorophenol (4-CP), 4-nitrophenol (4-NP) and 4-bromophenol (4-BP) in addition to unsubstituted phenol[1], at high concentrations of 1.44, 2.72 and 12.77 mM, respectively[2]. These compounds are common pollutants in soil, and 4-NP, in particular, is a priority pollutant listed by the U.S. Environmental Protection Agency (http://www.epa.gov/waterscience/methods/pollutants.htm). *A. chlorophenolicus* degrades these compounds efficiently as single growth substrates and in mixtures[3]. This bacterium can degrade 4-CP under mesophilic (28°C) and psychrophilic (5°C) conditions and during repeated fluctuations between these temperature extremes[4].

A. chlorophenolicus and many other members of the actinobacteria group are common residents of soil and have high tolerance to stressful conditions encountered in the soil environment. *A. chlorophenolicus* cells can survive after inoculation to non-sterile soil in a presumed dormant state[5], demonstrating an unusual stress tolerance that has also been reported for other *Arthrobacter* and related actinobacterial species[6]. Therefore, it would be interesting to understand the mechanisms underlying the ability of *A. chlorophenolicus* and related strains to survive and grow under these highly stressful conditions.

One adaptation mechanism to changing environmental parameters is to change the composition of fatty acids in the cell membrane. Previously, we studied adaptations of the *A. chlorophenolicus* cell membrane fatty acids to changes in concentrations of phenolic compounds and to temperature[2]. Clear effects on the *anteiso*:*iso* ratio of branched fatty acids were seen in response to increasing concentrations of phenols, and to an even higher extent in response to changes in growth temperature.

Our aim in this study was to examine mechanisms used by *A. chlorophenolicus* to adapt to growth on different phenolic compounds or temperatures at the protein level. We used a bottom-up, or "shotgun" mass spectrometry based proteomics approach[7,8] followed by label free quantitation and statistical analyses for detailed exploration of the *A. chlorophenolicus* proteome during growth on different phenolic compounds and at different temperatures. The shotgun proteomics approach is based on a coupling of multidimensional liquid chromatography with electrospray-tandem mass spectrometry (2D-LC-ES-MS/MS). This approach has many advantages over traditional 2-dimensional-polyacrylamide gel electrophoresis (2D PAGE-MS)[9,10]. With recent instrumental advances[11,12] 2D-LC-MS/MS methods can identify up to half of the predicted proteins (1000-3000 protein products) in isolates in a single experiment in a few days[13-17]. This method is highly dependent on parallel computational analyses of the predicted proteome sequence that is cleaved and fragmented *in silico* and compared with MS/MS spectra via search engines such as SEQUEST[7] and

Mascot[18]. The recent advent of the draft genome sequence of *A. chlorophenolicus* (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) enabled us to specifically study its proteome using the shotgun proteomics approach.

Our specific aims in the present investigation were to determine how *A. chlorophenolicus* adapts its proteome in response to stress conditions, such as temperature changes between 5 and 28°C and to growth on different phenolic substrates. In addition, we studied a mutant strain of *A. chlorophenolicus* (T99), harboring a non-functional hydroxyquinol 1,2-dioxygenase gene. This mutation disabled the cell's ability to grow on substituted phenols, but its growth on phenol was even better than the wild type[19]. Therefore, we also examined the proteome of mutant T99 and compared it to the wild type during growth on phenol to gain a better understanding of proteome adaptations to phenolic substrates.

Materials and Methods

Bacterial strains and culture conditions. *Arthrobacter chlorophenolicus* strain A6 was previously isolated from a soil slurry enriched with increasing concentrations of 4-chlorophenol (4-CP)[1]. *A. chlorophenolicus* mutant T99 contains a chloramphenicol resistance-conferring transposon inserted into a hydroxyquinol 1,2-dioxygenase gene, disabling its ability to grow on substituted phenols[19]. The cells were grown in GM minimal medium[1] supplemented with 4-nitrophenol (4-NP), 4-CP or phenol as previously described[1,2] at 28°C and 5°C. Mutant T99 was grown in cultures supplemented with 5 µg mL^{-1} chloramphenicol to select for transposon retaining cells. 4-NP and 4-CP were purchased from Sigma-Aldrich (Steinheim, Germany) and phenol from Merck (Darmstadt, Germany). The phenols were added from stock solutions to final concentrations of 100 µg mL^{-1} 4-NP, 150 µg mL^{-1} 4-CP and 400 µg mL^{-1} phenol.

Cell lysis and sample preparation for 2D-LC-MS/MS. Triplicate cultures from 8 different treatments were prepared. Cells in mid-log phase of growth (OD_{600}=0.08-0.3, depending on growth conditions as previously determined[1,2], were harvested by centrifugation at 6 000 × g for 20 min. They were washed once with 1.5 mL Tris-EDTA-buffer (TE), pH 7.6, at the same temperature used to grow the cells; i.e. 5 or 28°C. Cells were pelleted by centrifugation 5 min at 16 000 × g at 4°C and stored at -80°C until further treatment. Subsequently cell pellets were diluted in 1 mL TE for lysis using a FastPrep®-24 cell disrupter (MP Biomedicals, Solon, OH) in tubes containing 0.5 mL of 0.1 mm zirconium/silica beads (BioSpec Products, Inc., Bartlesville, OK) at a speed of 6.5 m/s four times, 45 s each. The lysate was centrifuged at 7000 × g for 20 min. and the supernatant was recovered. Concentrated trichloroacetic acid (TCA) was added to a volume of 10% and the samples were incubated at 4°C overnight. The resulting precipitate was harvested by centrifugation at 16 000 × g for 10 min., and washed with ice-cold methanol. After one additional centrifugation step, the precipitate was stored at -80°C until LC-MS analysis (see below). All samples were processed as follows: protein pellets were re-suspended in 6 M guanidine/10 mM dithiothreitol (DTT) in 50 mM Tris buffer (pH 7.6) and heated for 1 h at 60°C. The guanidine was then diluted 6-fold with 50 mM Tris buffer/10mM $CaCl_2$ (pH 7.6), proteins were digested into peptides with 1:100 (wt/wt) sequencing grade trypsin (Promega, Madison, WI), and cellular debris was removed by centrifugation (3000 × g for 10 min). Peptides were desalted off-line by C18 solid phase extraction (Waters, Milford, MA), concentrated (to 500 µL), filtered and aliquoted (150 µL per aliquot; entire aliquot used for each LC-MS analysis).

2D-LC-MS/MS. Two-dimensional nano LC-ES-MS/MS analysis of each biological replicate and sample type was carried out on an linear ion trap mass spectrometer (LTQ Thermo Fisher, San Jose, CA) as previously described[13,20]. In total eight different biological samples were analyzed with three biological replicates for each sample, except for two replicates for the samples from cultures grown on 4-NP at 28°C, due to insufficient protein

quantity in the third replicate for reliable analysis. All samples were run using the same methodology on the same LC-MS system. LC columns were thoroughly washed between sample sets to avoid carry over contamination or changed out entirely. Samples were loaded onto a 2-dimensional split phase column made of strong cation exchange (SCX) and reverse phase (RP). Samples (150 µL) were first loaded onto a 150 µm back column packed with of 4 cm of RP and and 4 cm of SCX. This back column was then connected to a 100 µm RP front column with an integrated nanospray tip that was packed with 15 cm of RP. The column system was placed into a nanospray source (Proxeon, Denmark), directly in front of the LTQ mass spectrometer. The LTQ was coupled to an Ultimate HPLC pump (LC Packings; a division of Dionex, San Francisco, CA), which had an initial flow rate of ~100 µL/min that was split precolumn to obtain a flow rate of ~300 nL/min at the nanospray tip. Samples were analyzed via two-dimensional liquid chromatography over 23 h by 11 consecutive increasing (0-500 mM) pulses of ammonium acetate salt. Each salt pulse was followed by a 2 h reverse phase gradient from 100% aqueous solvent (95% H_2O/ 5% ACN/ 0.1% formic acid) to 50% organic solvent (30% H_2O/ 70% ACN/ 0.1% formic acid). During the chromatographic separation the LTQ was operated in a data-dependent mode and under the direct control of the Xcalibur software (Thermo Fisher Scientific). The following parameters were applied to the LTQ MS analyses: nanospray voltage of 3.8 kV, heated capillary temperature of 200°C, and a full mass scan range of 400-1700. MS/MS spectra were acquired in data-dependent mode as follows: 5 MS/MS spectra were obtained following every full scan; 2 microscans were averaged for every full MS and MS/MS spectrum; a 3 m/z isolation width was employed; 35% collision energy was used for fragmentation, and dynamic exclusion was set to 1 with a duration of 3 min.

Resulting MS/MS spectra were searched using the DBDigger algorithm[22,23] with a database containing all the non-redundant predicted proteins from *A. chlorophenolicus* (5,286 entries) and 36 common contaminants (i.e. keratins and trypsin). The proteins in the database were designated according to their IMG gene object IDs from the draft genome sequence (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi). Cluster of Orthologous groups (COG) categories were also extracted from IMG and added to each protein entry. The output data files were then filtered and sorted with the DTASelect algorithm[24] using the following parameters: fully tryptic peptides only, with delCN of at least 0.08 and cross-correlation scores (Xcorrs) of at least 25 (+1), 30 (+2) and 45 (+3). At least two peptides had to be identified within the same run in order for a protein to be deemed identified. DTASelect output files were extracted for total proteins, peptides, spectra and sequence coverage for each protein. False positive levels were estimated via decoy database method with a reverse database appended on the end of the forward database[14]. Reverse entries were given a unique identifier and total peptides matching forward entries and reverse entries were calculated using the formula from Peng et al. [14]. False positive rates were calculated for 6 representative runs and resulted in an approximate false positive rate of 2.2-5.7% per run. It should be noted that only proteins with at least 5 total spectral counts were used for quantitation (discussed below) that further reduced the false positive rate for quantified proteins.

Protein abundance was estimated using a semi-quantitative, label free approach by counting the number of MS spectra for each protein in an individual run[25,26]. Spectral counts from each growth condition and LC-MS analysis were extracted into a single worksheet. Proteins with 5 or fewer total spectral counts across all samples were excluded from the statistical analysis due to low reliability of the abundance of these proteins. The dataset was analyzed using the Poisson regression model that is commonly used for count data based on the assumption that the data have a Poisson distribution, such as frequently encountered when counting a number of events[27]. The Poisson regression model assumes the logarithm of its

expected value can be modeled by a linear combination of the independent variables. In our analysis, we used spectral counts as the outcome and growth condition as the independent variable. To make the spectral counts comparable across different experiments, we normalized the spectral counts for a protein to the total spectral counts in a specific experiment. In Poisson regression, this is handled by adding the logarithm of total spectral count as an independent variable with a fixed coefficient of 1. The *p* values generated by the model were further adjusted using the Benjamini and Hochberg correction to account for multiple comparisons[28]. An adjusted *p* value of 0.01 (i.e. 1% False Discovery Rate) was used to select proteins that were differentially expressed between selected groups.

Classification of proteins & visualization. In order to classify the proteins identified via LC-MS analysis based on enzymatic function, we used the enzyme profiles (as of July 2006) provided by PRIAM[29]. For all predicted protein sequences, a RPS-BLAST[30] against the enzyme profiles was performed and the results were parsed according to the cutoffs given by PRIAM. Identified enzymes were then mapped to KEGG pathway maps[31] based on whether their abundances were found to be changed in the respective comparison.

Results and Discussion

Using the shotgun proteomics approach we could identify between 1645 and 2074 proteins per sample and 3773 non-redundant proteins in total from samples from all growth states, including growth at different temperatures and on different phenolic compounds, and a mutant with an inactivated 4-CP degradation pathway during growth on phenol (Supplementary Table 1). The number of proteins, peptides and spectra from each growth condition and biological replicate are shown in Table 1. The draft genome of *A. chlorophenolicus* at the time of screening contained 5286 potential protein coding genes and since not all of these are expressed at any given time (i.e. typically only a third are expressed in any given growth condition due to gene regulation), this is excellent coverage of the predicted proteome of this

microorganism. *A. chlorophenolicus* was found to have large differences in its proteome depending on the growth temperature or growth substrate. Some of the most abundant proteins regardless of growth substrate or temperature were e.g. a monooxygenase (2500143301) (previously reported as CphC-I[19]), a chaperonin GroEL (2500146253) and a chaperone protein DnaK (2500143310) as well as a Succinyl-CoA synthetase (2500145054). The most abundant proteins in each growth condition are listed in Table 2.

Proteins from each growth condition were grouped in clusters of orthologous (COG) categories, although please note our disclaimer that these groupings were only made using a draft genome and they are likely to change with a completely finished genome. Still, it enabled us to make comparisons between the growth conditions. At this high level of analysis, however, no major differences were seen between the different growth conditions though minor differences were noted (Supplementary Table 2). The dominant COG categories included hypothetical or conserved hypothetical proteins, amino acid transport and metabolism, translation and energy production. While proteins were found in all categories, some including cytoskeleton, motility, chromatin structure and RNA processing and modification had only 1-4 protein identifications per growth state. This is not surprising since some of these functions are rarely seen or not used at all in bacteria. A “core” proteome was extracted that contained only those proteins found in all growth conditions and biological replicates. In total 738 proteins were found in all samples and biological replicates (Supplementary Table 3), interestingly these had the same distribution of COG categories as the entire set (Supplemental Table 2). Proteins found in this subset included many absolutely necessary for growth including most of the ribosomal and translation proteins, transcription proteins, chaperones, oxidative phosphorylation proteins, and other core metabolic pathways. Interestingly, many hypothetical or conserved hypothetical proteins were found in the core proteome suggesting a critical need for these proteins.

An overall comparison of 1678 genes showing significantly different expression profiles in

comparisons between at least two growth states is shown in Figure 1. The main findings from these comparisons can be summarized as follows: 1) The major differences in protein expression depend on the substrate and phenol is very different from the other two substrates. 2) Among samples grown on 4-NP and 4-CP, the temperature effect is stronger than the substrate effect. 3) In comparisons of the wild type and mutant strains grown on phenol, the temperature effect is stronger than the mutation effect. 4) Cultures grown on phenol are more sensitive to temperature changes than those grown on 4-NP and 4-CP. The different individual comparisons are discussed in more detail in the following sections. Due to the large number of significant changes in levels of many different proteins it is difficult to dissect the complex interplay between regulation of their expression and functional significance. However, through deep exploration of the proteomes we found indications of interesting and sometimes unexpected mechanisms for stress adaptations in this species and some of these are highlighted in the following sections.

Differences in protein expression at 28°C vs. 5°C

Cold shock proteins

In a preliminary experiment, 2D-polyacrylamide gel electrophoresis revealed several differences in the proteomes of *A. chlorophenolicus* cells grown at 5°C compared to cells grown at 28°C, irrespective of the growth substrate (data not shown). One spot was prominent in 5°C-cultures but not present or faint at 28°C. This spot was *de novo* sequenced and identified as a putative DNA-binding cold-shock protein, homologous to the CapA protein found in *A. globiformis* SI55[32].

Subsequently, we investigated the same temperature growth conditions using the shotgun proteomics approach. Hundreds of proteins differentiated between the two growth temperatures. The same putative cold-shock protein (2500145448) that we found using the 2D gel approach was identified as consistently more highly expressed at 5°C in the shotgun proteomics data. Five other homologues to this protein were also expressed in all cultures,

however their expression was not consistently higher at 5°C compared to 28°C, indicating that these proteins do not offer cold protection only, but have other stress-protective roles as well, as previously suggested[33].

Protein variation at 5 and 28°C independent of growth substrate

Only 25 proteins were consistently significantly differently expressed at the two temperatures regardless of growth substrate (Table 3). Several of these were chaperonins and ribosomal proteins that are known to respond to temperature stress. Some unknown proteins also differed in response to temperature, reflecting the need for future research in the area of temperature adaptations. Only 5 of the 25 proteins mentioned consistently changed in the same direction in response to temperature. These were a putative monooxygenase (2500143405), an alanine dehydrogenase (2500145508), a hypothetical protein (2500145780) (re-named unknown), a formaldehyde dehydrogenase (2500143393), and a putative cold-shock DNA-binding domain protein (2500145448). Of these, all but the putative monooxygenase and the formaldehyde dehydrogenase increased at 5°C compared to 28°C. We can only speculate about the roles of these proteins in temperature adaptations at this stage. Alanine dehydrogenase has previously been reported as associated with growth at low oxygen levels and accompanying cell dormancy as well as starvation, nitrogen starvation or salt stress[34-37], and is necessary for sporulation in *Bacillus subtilis*[38]. However, to our knowledge it has not been associated with cold growth until now.

Protein variation at 5 and 28°C depending on substrate

Different types and amounts of proteins varied at the two temperatures depending on the growth substrate: 4-CP, 109 proteins; 4-NP, 222 proteins; phenol, 527 proteins. Some of these are listed in Table 4 and discussed further below. In particular, cultures grown on 4-NP reacted differently to changes in temperature compared to those grown on the other two phenolic compounds. Among the proteins up-regulated at 5°C compared to 28°C in 4-CP and/or phenol cultures but not in 4-NP cultures was an AAA ATPase (ATPases Associated

with diverse cellular Activities) (2500143793) having chaperone-like functions, and a peptidylprolyl isomerase (2500147455) which facilitates protein folding. Such proteins may help to facilitate replication, transcription and translation processes that have reduced efficiency at low temperatures. In addition, a GplX protein (2500145962) was more highly expressed at 5°C than 28°C in 4-CP and phenol-grown cultures but not in 4-NP-grown cultures. The GplX protein is involved in glycerol metabolism, indicating that glycerol might play a role in bacterial cold adaptation at least during some conditions in *A. chlorophenolicus*, as in other organisms, e.g. *Saccharomyces cerevisiae*[39], and as for other kinds of stresses, e.g. osmotic stress[40].

One of the proteins that had a lower abundance during cold growth on 4-CP and phenol was the uncharacterized conserved protein YceI (2500145348). This protein has probable importance for isoprenoid quinone metabolism and for controlling oxidative stress and gene regulation[41]. The role of this protein in adaptation to changes in temperature is currently difficult to predict, if it is not part of a general stress response. It is also discussed in the context of growth substrate adaptations below.

Temperature adaptation of membrane fatty acid composition

Previously, we found that the *anteiso*:*iso* ratio of the *A. chlorophenolicus* membrane fatty acids decreased in response to high temperatures and increasing concentrations of phenolic compounds, the extent depending on the nature of the phenolic compound[2]. The observed relative increase in expression of a dihydroxy-acid dehydratase (2500143431) in response to high temperatures in this study would thus be consistent with the results of the previous study. This indicates a role of this protein in production of *iso* branched fatty acids. The mechanism that Gram positive bacteria use to control their *anteiso/iso* ratio is not known, but this could be based on the ratio between the precursors valine and leucine for *iso*, and isoleucine for *anteiso*. The results of this study therefore provide a first indication of a mechanism for Gram positive bacteria to control their membrane fluidity, through expression of e.g. the protein

dihydroxy-acid dehydratase and subsequent production of precursors for *iso*-branched fatty acids.

Differences in protein expression in response to phenolic substrate. To study differences in protein expression due to the phenolic growth substrate, samples from the temperature growth optimum of 28°C were compared. In general, the protein expression pattern was more similar for growth on the substituted phenols, compared to phenol. In the comparison between 4-CP and 4-NP 101 proteins were significantly differently expressed, whereas 255 vs. 203 differed between 4-NP and phenol vs. 4-CP and phenol, respectively. In the latter two lists, 111 protein identities were shared that did not differ between 4-CP and 4-NP cultures. This clearly illustrates that 4-CP and 4-NP cultures share a more similar protein expression pattern compared to phenol-grown cultures. This could be a result of common degradation routes for 4-CP and 4-NP (Figure 2) compared to phenol, as suggested in a previous study[3], and/or due to the higher toxicity of 4-NP and 4-CP compared to phenol. All proteins discussed below are listed in Table 5.

<u>Proteins more abundant during growth on substituted phenols compared to phenol</u>

The YceI protein mentioned above was much more abundant in cultures grown on 4-CP and 4-NP compared to those grown on phenol. This protein was also shown to increase in abundance in *Delftia acidovorans* after addition of a chlorophenoxy herbicide as a growth substrate[42]. It may be a stress response towards halogenated aromatic hydrocarbons as members of this family are known to control oxidative stress[41]. The probable role of YceI in quinone metabolism may also have importance for the preference of *para*-substituted phenols, having a quinone-like structure, in *A. chlorophenolicus*[1].

We found that a monooxygenase previously reported as CphC-I[19] (2500143301), corresponding to the oxygenase component of a monooxygenase and part of the proposed 4-chlorophenol degradation pathway, was more highly expressed in 4-CP and 4-NP-grown cultures than in phenol-grown cultures. This makes sense considering its predicted role in

metabolism of 4-CP and 4-NP, but not phenol. In addition, this protein had the highest spectral counts in all samples regardless of growth substrate or temperature.

In addition, several ATP synthases (ATPases) (2500146385 and 2500146382; 2500143793, 2500146381 and 2500146380) were expressed at higher levels during growth on substituted phenols. Previous studies of pseudomonads have shown that the ATP concentration decreases in response to exposure to lipophilic hydrocarbons[43,44]. In addition, it is known that phenolic compounds uncouple cellular respiration[45,46], which could cause inefficient ATP synthesis and result in low ATP concentrations. Therefore, the increase in expression of ATPases in the presence of 4-CP and 4-NP, relative to phenol, may be an effort to compensate for the decreased ATP concentrations. Such indications of compensation for low levels of ATP have not been reported previously to our knowledge.

Some chaperone and chaperonin proteins were more abundant during growth on substituted phenols compared to phenol: a chaperone DnaK (2500143310) and its co-chaperone, DnaJ (2500144536) that are part of the Hsp70 heat-shock system, involved in protein folding and renaturation after stress; GroEL proteins (2500146253, 2500145671, 2500146254) and a SufBD protein (2500145932), that is an important cofactor for numerous proteins involved in the SUF system operating under e.g. oxidative stress, which can occur when cells are subjected to lipophilic hydrocarbons as in this study.

Proteins that were specifically higher in abundance in 4-NP grown cultures included those with possible roles in nitrogen metabolite repression, i.e. NmrA family protein (2500143164)[47] and the dihydroxy-acid dehydratase (2500143431) that was discussed above when comparing growth temperatures, strengthening our hypothesis that this enzyme is associated with *iso* branched fatty acid synthesis, since we previously found the highest levels of *iso* branched fatty acids when *A. chlorophenolicus* was grown on 4-NP[2].

Proteins with higher abundances in phenol grown cultures

Five additional putative monooxygenases in two clusters on separate contigs (2500143405, 2500143406, 2500143407; and 2500144277, 2500144278) were 6-30 times more abundant in phenol-grown wild type cultures than 4-NP or 4-CP-grown cultures. In addition, CphA-II (2500143295), the second hydroxyquinol 1,2-dioxygenase in the previously reported 4-CP degradation pathway[19] (Figure 2), was at levels almost four times higher for phenol grown cultures than for those grown on 4-CP or 4-NP. This suggests that CphA-II is likely involved in degradation pathways of *Arthrobacter chlorophenolicus* cells grown on all three phenolic substrates.

Two glycerol kinases (2500146342, 2500146341) and a glycerone kinase (2500143798), were much more abundant in phenol cultures than 4-NP or 4-CP cultures, suggesting that glycerol is used as a reserve carbon source when *A. chlorophenolicus* is growing on phenol. This data suggests that phenol is not a sufficient energy source on its own, explaining the slow and inconsistent growth seen on this substrate compared to the other phenolic compounds[3].

Another indication of insufficient energy during growth on phenol is the elevated levels of key enzymes in the glyoxylate bypass during growth on this compound; i.e. isocitrate lyase (2500144161) and malate synthase proteins (2500143966, 2500144162). The glyoxylate bypass is a shunt in the tricarboxylic acid (TCA) cycle, permitting the use of fatty acids or acetate as carbon sources to provide intermediates in the TCA cycle (Figure 3a). In addition, the T99 mutant with a disrupted 4-chlorophenol degradation operon, also had much higher levels of the glyoxylate bypass proteins than the wild type strain when both were grown on phenol (Figure 3b, Table 7). Since the mutant is known to grow better on phenol than the wild type strain[3], these data suggest that the glyoxylate bypass is a beneficial energy yielding reaction during growth on phenol.

<u>Differences in uptake/transport mechanisms for the different phenolics</u>

Several uptake and transport-associated proteins also differed in expression levels between phenol and the substituted phenols suggesting that different uptake mechansisms are used for

transport of these two classes of phenolic compounds. We previously found that when 4-CP, 4-NP and phenol were added together as a mixture, 4-NP was degraded first, followed by 4-CP and then phenol[3]. In addition, phenol degradation did not begin until 4-CP was almost depleted indicating some sort of interaction between the compounds. Examples of transporter related proteins that were only detected during growth on phenol, but not substituted phenols were: extracellular solute-binding proteins (2500144279, 2500145620, 2500143411), an ABC transporter related protein (2500145130) and a putative ABC-type sulfonate transport system protein (2500145128). In contrast, a different extracellular solute-binding protein (2500144130) was more highly expressed during growth on the substituted phenolic compounds. There were also two periplasmic binding proteins and another extracellular solute-binding protein family 1 (2500145912, 2500144463, 2500144131), mainly expressed in 4-NP cultures, suggesting special transport proteins for this compound. Previous kinetic analyses[3] indicate that 4-NP and 4-CP are in fact degraded simultaneously, but 4-NP has a faster rate of transport into the cell and this could be due to differences in their pKa values, should the phenolate ion be the preferred uptake substrate in both cases. However, study of the proteomes suggest that the situation is more complex, involving different uptake mechanisms depending on the substrate. The different expression of several ABC transporter proteins between wild type and the T99 mutant strain is interesting, since our previous study indicated transport competition between 4-CP and phenol[3].

Unknown proteins

Several unknown or hypothetical (re-named unknown) proteins differed between 4-CP vs. phenol, 4-NP vs. 4-CP and 4-NP vs. phenol-grown cultures; 16, 10 and 23, respectively. None of these proteins was significantly differently expressed across all growth substrates. Eight of these were different when comparing 4-CP and 4-NP vs. phenol, but not between 4-CP and 4-NP grown cultures, an additional indication of a similar behaviour of *A. chlorophenolicus* on substituted phenols compared to phenol. Five were more abundant in phenol-grown cultures

(2500145574, 2500146173, 2500147824, 2500146860, 2500143408) and one of these (2500143408) was almost 10 times more abundant in phenol cultures. There was also a protein that was 4 times more abundant in 4-CP cultures than in the other cultures (2500144629). The large differences we observed in expression of these uncharacterized proteins reflects the need for further research in the area of adaptations to aromatic compounds and their degradation pathways.

Wild type versus mutant

When comparing the proteomes of wild type and mutant T99 strains during growth on phenol at 28°C, 453 proteins differed significantly in their abundances (see Table 6 for a partial list). Many of the proteins found to a higher extent in the wild type strain compared to the mutant were those in the chlorophenol degradation pathway as expected, since this pathway is disrupted in the mutant[19] (Table 6). The putative monooxygenases in two clusters on separate contigs (2500143405, 2500143406, 2500143407; and 2500144277, 2500144278) that were abundant in phenol-grown wild type cultures (see discussion above) were not detected in the mutant. Downstream of the disrupted *cphA-I* gene in the mutant is an ORF, *cphX* (not annotated in the draft genome), similar to an ABC transporter and this protein was also not detected in the mutant. This is also true of an ABC transporter related protein (2500145130), a putative ABC-type sulfonate transport system (2500145128) and two extracellular solute-binding proteins, family 3 and 5 (2500143411, 2500144279). Thus, the mutation in T99 affects other genes apart from the knocked-out gene, and the mutant might even degrade phenol differently than the wild type.

Interestingly, an extracellular solute-binding protein family 1 (2500143192) was more highly expressed in the mutant compared to the wild type strain and an ABC-transporter related protein (2500143447) was only found in the mutant. These findings suggest that transport of phenol into the cell could be occurring differently in the mutant compared to the

wild type strain. Since the mutant grows better on phenol than the wild type strain a more efficient transport of phenol into the cell could be an alternative explanation for this finding. Although we do not know the pathway used for phenol degradation in the mutant (or wild type) there are several clues obtained from studying the mutant. The hydroxyquinol 1,2-dioxygenase, CphA-I, is evidently redundant for phenol degradation, since the mutant strain can grow better on this compound than the wild-type strain without producing this protein. Although levels of the second hydroxyquinol 1,2 dioxygenase, expressed by *cphA-II* in the 4-chlorophenol degradation gene cluster, were lower in the mutant than in the wild type it was still expressed at relatively high levels. Since the mutant is unable to grow on substituted phenols, the CphA-II enzyme is not sufficient for that process (Table 6). The two CphA enzymes are sufficiently different at the protein level to be distinguished by high resolution MS, therefore, these peptides were not misidentified. Differences in KEGG metabolic pathways indicated by differences in protein expression between wild type and mutant are shown in Table 7. Besides the differences already mentioned in glyoxylate and dicarboxylate metabolism (see also Fig. 3) and benzoate degradation that encompasses some of the proteins involved in 4-CP degradation, are some major differences in purine and pyrimidine metabolism. Also, some proteins involved in biosynthesis of valine, leucine and isoleucine are higher in the mutant than the wild type.

Conclusion. In summary, the results of this study provide a glimpse of the complicated processes behind adaptation to growth in low temperature as well as to growth on different phenolic substrates. Many adaptation mechanisms are employed by *A. chlorophenolicus* in response to temperature stress and phenolic substrates, and often the same proteins are expressed as a response towards both kinds of stresses. Increasing studies show that stress-induced proteins are often the same for many different kinds of stresses, and hence maybe their names should be re-evaluated, as in the case of e.g. cold-shock or heat-shock proteins.

These data have also revealed an unexpectedly complicated machinery of phenol degradation pathway(s) in this organism, a question that requires further research to solve.

Acknowledgements. We want to acknowledge Professor Lawrence P. Wackett, University of Minnesota, for valuable discussions about possible phenol degradation pathways in *A. chlorophenolicus* and Dr. Jörn Kalinowski (Institute for Genome Research and Systems Biology, Center for Biotechnology, Bielefeld University, Bielefeld, Germany) for assistance in identification of up regulated proteins in different growth conditions and assigning these to KEGG maps (Figure 3). The ORNL part of this research was sponsored by a Work For Others with the Swedish University of Agricultural Sciences, Uppsala, Sweden. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. This study was supported by the Department of Microbiology, Swedish University of Agricultural Sciences, Uppsala, Sweden and in part by the U.S. Department of Energy Contract No. DE-AC02-05CH11231 with Lawrence Berkeley National Laboratory. The draft *A. chlorophenolicus* genome sequence was provided through a Community Sequencing Project grant to JKJ from the DOE Joint Genome Institute, Walnut Creek, CA, USA.

Supporting Information Available

Supplementary Tables 1-3 are available free at http://pubs.acs.org.

Supplementary Table 1 contains all proteins identified across all growth conditions with spectral counts for each identification.

Supplementary Table 2 illustrates a comparison of the total proteins found from each COG group for each growth condition as well as a “core” proteome.

Supplementary Table 3 contains all proteins found in all growth conditions and biological replicates, the “core” proteome.

References

(1) Westerberg, K., Elväng, A. M., Stackebrandt, E. and Jansson, J. K. (2000) *Int. J. Syst. Evol. Microbiol.* **50**, 2083-2092.
(2) Unell, M., Kabelitz, N., Jansson, J. K. and Heipieper, H. J. (2007) *FEMS Microbiol. Lett.* **266**, 138-143.
(3) Unell, M., Nordin, K., Jernberg, C., Stenstrom, J. and Jansson, J. K. (2008) *Biodegradation*. **19**, 495-505.
(4) Backman, A. and Jansson, J. K. (2004) *Microb. Ecol.* **48**, 246-53.
(5) Backman, A., Maraha, N. and Jansson, J. K. (2004) *Appl. Environ. Microbiol.* **70**, 2952-2958.
(6) Mongodin, E. F., Shapir, N., Daugherty, S. C., Deboy, R. T., Emerson, J. B., Shvartzbeyn, A., Radune, D., Vamathevan, J., Riggs, F., Grinberg, V., Khouri, H., Wackett, L. P., Nelson, K. E. and Sadowsky, M. J. (2006) *Public Library of Science Genetics*. **2**, 2094-2106.
(7) Eng, J. K., McCormack, A. L. and Yates, J. R. (1994) *J. Am. Soc. Mass Spectrom.* **5**, 976-989.
(8) Washburn, M. P., Wolters, D. and Yates, J. R. (2001) *Nat. Biotechnol.* **19**, 242-247.
(9) Liu, H. B., Lin, D. Y. and Yates, J. R. (2002) *Biotechniques*. **32**, 898-+.
(10) Peng, J. M. and Gygi, S. P. (2001) *J. Mass Spectrom.* **36**, 1083-1091.
(11) Hu, Q. Z., Noll, R. J., Li, H. Y., Makarov, A., Hardman, M. and Cooks, R. G. (2005) *J. Mass Spectrom.* **40**, 430-443.
(12) Schwartz, J. C., Senko, M. W. and Syka, J. E. P. (2002) *J. Am. Soc. Mass Spectrom.* **13**, 659-669.
(13) Brown, S. D., Thompson, M. R., VerBerkmoes, N. C., Chourey, K., Shah, M., Zhou, J. Z., Hettich, R. L. and Thompson, D. K. (2006) *Mol. Cell. Prot.* **5**, 1054-1071.
(14) Peng, J. M., Elias, J. E., Thoreen, C. C., Licklider, L. J. and Gygi, S. P. (2003) *J Prot Res.* **2**, 43-50.
(15) VerBerkmoes, N. C., Shah, M. B., Lankford, P. K., Pelletier, D. A., Strader, M. B., Tabb, D. L., McDonald, W. H., Barton, J. W., Hurst, G. B., Hauser, L., Davison, B. H., Beatty, J. T., Harwood, C. S., Tabita, F. R., Hettich, R. L. and Larimer, F. W. (2006) *J Prot Res.* **5**, 287-298.
(16) Zhu, W. H., Reich, C. I., Olsen, G. J., Giometti, C. S. and Yates, J. R. (2004) *J Prot Res*. **3**, 538-548.
(17) Kolker, E., Picone, A. F., Galperin, M. Y., Romine, M. F., Higdon, R., Makarova, K. S., Kolker, N., Anderson, G. A., Qiu, X. Y., Auberry, K. J., Babnigg, G., Beliaev, A. S., Edlefsen, P., Elias, D. A., Gorby, Y. A., Holzman, T., Klappenbach, J. A., Konstantinidis, K. T., Land, M. L., Lipton, M. S., McCue, L. A., Monroe, M., Pasa-Tolic, L., Pinchuk, G., Purvine, S., Serres, M. H., Tsapin, S., Zakrajsek, B. A., Zhou, J. H., Larimer, F. W., Lawrence, C. E., Riley, M., Collart, F. R., Yates, J. R., Smith, R. D., Giometti, C. S., Nealson, K. H., Fredrickson, J. K. and Tiedje, J. M. (2005) *Proceedings of the National Academy of Sciences of the United States of America*. **102**, 2099-2104.
(18) Perkins, D. N., Pappin, D. J. C., Creasy, D. M. and Cottrell, J. S. (1999) *Electrophoresis*. **20**, 3551-3567.
(19) Nordin, K., Unell, M. and Jansson, J. K. (2005) *Appl. Environ. Microbiol.* **71**, 6538-6544.
(20) Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., Shah, M., Hettich, R. L. and Banfield, J. F. (2005) *Science*. **308**, 1915-1920.
(21) McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J. and Yates, J. R. (2002) *International Journal of Mass Spectrometry*. **219**, 245-251.
(22) Tabb, D. L., Narasimhan, C., Strader, M. B. and Hettich, R. L. (2005) *Anal. Chem.* **77**, 2464-2474.
(23) Narasimhan, C., Tabb, D. L., VerBerkmoes, N. C., Thompson, M. R., Hettich, R. L. and Uberbacher, E. C. (2005) *Anal. Chem.* **77**, 7581-7593.
(24) Tabb, D. L., McDonald, W. H. and Yates, J. R. (2002) *J Prot Res.* **1**, 21-26.
(25) Liu, H. B., Sadygov, R. G. and Yates, J. R. (2004) *Anal. Chem.* **76**, 4193-4201.
(26) Zhang, B., VerBerkmoes, N. C., Langston, M. A., Uberbacher, E., Hettich, R. L. and Samatova, N. F. (2006) *J Prot Res*. **5**, 2909-2918.
(27) Dobson. (2001)*An Introduction to Generalized Linear Models*, 2 ed., London.
(28) Benjamini, Y. and Hochberg, Y. (1995) *J. R. Statist. Soc. B* **57**, 289-300.
(29) Claudel-Renard, C., Chevalet, C., Faraut, T. and Kahn, D. (2003) *Nucleic Acids Res*. **31**, 6633-6639.
(30) Marchler-Bauer, A., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Lu, S., Marchler, G. H., Mullokandov, M., Song, J. S., Tasneem, A., Thanki, N., Yamashita, R. A., Zhang, D., Zhang, N. and Bryant, S. H. (2009) *Nucleic Acids Res*. **37**, D205-210.
(31) Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) *Nucleic Acids Res*. **34**, D354-357.
(32) Berger, F., Normand, P. and Potier, P. (1997) *J. Bacteriol.* **179**, 5670-5676.
(33) Givskov, M., Eberl, L., Moller, S., Poulsen, L. K. and Molin, S. (1994) *J. Bacteriol.* **176**, 7-14.
(34) Betts, J. C., Lukey, P. T., Robb, L. C., McAdam, R. A. and Duncan, K. (2002) *Molecular Microbiology*. **43**, 717-731.
(35) Lahmi, R., Sendersky, E., Perelman, A., Hagemann, M., Forchhammer, K. and Schwarz, R. (2006) *J. Bacteriol.* **188**, 5258-5265.
(36) Duche, O., Tremoulet, F., Glaser, P. and Labadie, J. (2002) *Appl. Environ. Microbiol*. **68**, 1491-1498.
(37) Hutter, B. and Dick, T. (1998) *FEMS Microbiol. Lett.* **167**, 7-11.
(38) Siranosian, K. J., Ireton, K. and Grossman, A. D. (1993) *J. Bacteriol.* **175**, 6789-6796.
(39) Aguilera, J., Randez-Gil, F. and Prieto, J. A. (2007) *FEMS Microbiol Rev*. **31**, 327-341.
(40) Beales, N. (2004) *Comprehensive Reviews in Food Science and Food Safety*. **3**, 1-20.
(41) Handa, N., Terada, T., Doi-Katayama, Y., Hirota, H., Tame, J. R. H., Park, S. Y., Kuramitsu, S., Shirouzu, M. and Yokoyama, S. (2005) *Protein Science*. **14**, 1004-1010.
(42) Benndorf, D., Davidson, I. and Babel, W. (2004) *Microbiology - Society for General Microbiology*. **150**, 1005-1014.

(43) Krayl, M., Benndorf, D., Loffhagen, N. and Babel, W. (2003) *Proteomics*. **3**, 1544-1552.
(44) Loffhagen, N., Hartig, C. and Babel, W. (2001) *Ecotoxicology and Environmental Safety*. **50**, 65-71.
(45) Escher, B. I., Snozzi, M. and Schwarzenbach, R. P. (1996) *Environ Sci Technol*. **30**, 3071-3079.
(46) Terada, H. (1990) *Environmental Health Perspectives*. **87**, 213-8.
(47) Stammers, D. K., Ren, J., Leslie, K., Nichols, C. E., Lamb, H. K., Cocklin, S., Dodds, A. and Hawkins, A. R. (2001) *Embo Journal*. **20**, 6619-6626.

Table of Contents (TOC) Synopsis

Arthrobacter chlorophenolicus is a psychrotrophic bacterium capable of growth on different phenolic compounds at high concentrations and at different temperature extremes. The *A. chlorophenolicus* proteome revealed hundreds of differentially expressed proteins, providing clues to stress adaptation mechanisms used by this microorganism. Study of the wild type and a mutant strain revealed insight into degradation pathway(s) for phenol and substituted phenolic compounds and metabolic shifts during growth on these compounds.

Table of

Table 1. The total number of proteins, peptides and spectra from each growth condition and biological replicate.

Growth condition	Proteins	Peptide IDs	Spectra
Wild type			
150 ppm 4-CP, 28°C	1856	14355	24213
replicate	1985	17476	29275
replicate	1831	16056	26946
100 ppm 4-NP, 28°C	1991	16298	30930
replicate	1711	14385	32640
400 ppm phenol, 28°C	1916	15304	28279
replicate	1841	14132	28018
replicate	1837	13192	24966
150 ppm 4-CP, 5°C	1988	17022	30280
replicate	1943	15537	28419
replicate	1804	14994	27988
100 ppm 4-NP, 5°C	2074	16236	32830
replicate	2000	16149	28191
replicate	1956	16564	28333
400 ppm phenol, 5°C	2052	18124	32659
replicate	2050	17924	32909
replicate	2004	16136	26593
Mutant T99			
400 ppm phenol, 28°C	1645	11330	27911
replicate	1650	13623	22713
replicate	2121	17130	29957
400 ppm phenol, 5°C	1777	13777	31246
replicate	1980	14682	30014
replicate	1955	15941	31116

Table 2. The most abundant proteins (spectral counts) in each growth condition, reported as averages of three replicates.

Protein description	COG	4-CP 28	4-CP 5	4-NP 28	4-NP 5	Phenol 28	Phenol 5
Monooxygenase (cphC-I)[a]	Q	621	498	604	768	470	667
YceI family protein	S	422	244	423	409	*122*	*82*
Chaperonin GroEL	O	384	624	736	443	456	202
Chaperonin Cpn60/TCP-1	O	371	465	506	348	311	*163*
Chaperone protein DnaK	O	318	364	456	309	257	258
Succinyl-CoA synthetase, beta subunit	C	244	352	394	353	317	324
Chaperonin Cpn60/TCP-1	O	243	288	312	240	*195*	*138*
Extracellular solute-binding protein family 5	E	243	225	314	326	252	380
Elongation factor Tu domain protein	J	189	216	*175*	*119*	*117*	*164*
Glyceraldehyde-3-phosphate dehydrogenase, type I	G	186	223	228	223	206	218

Maximum standard variation = 82%. Values in normal font are among the 10 most abundant for a given growth condition; whereas those in italics are still abundant, but not among the top 10 for all growth conditions and are shown for comparison. COG categories found at http://www.ncbi.nlm.nih.gov/COG/grace/fiew.cgi.

[a] Reported in [19].

Table 3. Proteins differing significantly (spectral counts) between temperatures regardless of growth substrate.

Protein description	COG	4-CP 28	4-CP 5	4-NP 28	4-NP 5	Phenol 28	Phenol 5
Monooxygenase (*cphC-I*)[a]	Q	621	498	604	768	470	667
chaperonin GroEL	O	384	624	736	443	456	202
chaperonin Cpn60/TCP-1 (GroEL)	O	371	465	506	348	311	163
conserved unknown protein	R	183	148	50	133	58	145
ATP synthase F1, alpha subunit	C	178	131	218	140	96	169
ribosomal protein L20	J	141	223	273	168	239	142
carbohydrate kinase FGGY	C	135	111	101	197	248	206
protein of unknown function DUF1486	none	110	215	219	136	125	200
chaperonin Cpn10	O	103	173	156	82	116	37
ribosomal protein L4/L1e	J	82	170	191	98	108	86
ribosomal protein L10	J	82	141	206	94	106	78
2-oxoglutarate dehydrogenase E2 component	C	130	100	173	124	87	139
Phosphoglycerate kinase	G	87	129	98	127	142	112
flavin reductase domain protein FMN-binding (*cphB*)[a]	R	64	29	40	67	64	24
Hydroxyquinol 1,2-dioxygenase (*cphA-II*)[a]	Q	51	117	56	134	184	133
ribosomal protein L18	J	47	75	95	54	38	64
ribosomal protein L11	J	37	62	90	48	64	46
FAD dependent oxidoreductase	C	22	7	4	20	11	29
putative cold-shock DNA-binding domain protein	K	8	43	23	51	19	56
putative monooxygenase	C	7	0	7	0	62	8
alanine dehydrogenase	E	5	35	4	27	6	94
CBS domain containing protein	R	1	8	14	4	13	0
unknown protein	none	0	17	0	16	0	49
Formaldehyde dehydrogenase glutathione-independent	E, R	20	5	28	10	15	2
transcriptional regulator, PadR-like family	K	3	0	0	4	1	10

>501
401-500
301-400
201-300
121-200
61-120
41-60
21-40
11-20
6-10
0-5

[a]Genes in previously reported 4-chlorophenol degradation cluster[19].

Table 4. Differences in protein expression (spectral counts) in response to growth temperature.

Protein description	COG	4-CP 28	4-CP 5	4-NP 28	4-NP 5	Phenol 28	Phenol 5	I.D.
YceI family protein	S	422	244	423	409	122	82	2
ATPase AAA-2 domain protein	O	90	199	94	73	58	187	2
Putative cold-shock DNA-binding domain protein	K	82	149	145	119	129	190	2
GlpX family protein	G	66	136	71	68	58	133	2
Putative cold-shock DNA-binding domain protein	K	61	138	122	90	102	146	2
Ribosomal protein L14	J	45	79	83	53	40	52	2, 3
Peptidylprolyl isomerase	O	43	68	53	50	48	84	1, 2
Dihydroxy-acid dehydratase	E,G	25	17	43	13	9	9	1, 3
Aldo/keto reductase	R	13	39	14	13	9	25	2
Conserved hypothetical protein	none	6	21	13	15	15	24	2, 3
Protein of unknown function DUF1684	none	5	0	4	1	37	4	2

I.D. = Numbers in last column indicate insignificant differences between samples from cultures grown on the following phenolic compound: 1) 4-CP, 2) 4-NP, 3) phenol. The color legend is the same as in Table 3.

Table 5. Differences in protein expression (spectral counts) in response to growth substrate.

Protein description	COG	4-CP	4-NP	Phenol	I.D.
YceI family protein	S	422	423	122	1
ATP synthase F1, alpha subunit	C	178	218	96	1
ATP synthase F1, beta subunit	C	156	137	76	
ATPase AAA-2 domain protein	O	90	94	58	1
H+transporting two-sector ATPase alpha/beta subunit domain protein	C	64	25	8	
H+transporting two-sector ATPase delta/epsilon subunit	C	17	20	2	1
Putative cold-shock DNA-binding domain protein	K	82	145	129	2
Putative cold-shock DNA-binding domain protein	K	61	122	102	2
Putative cold-shock DNA-binding domain protein	K	8	23	19	2
Putative cold-shock DNA-binding domain protein	K	42	41	73	1
Putative cold-shock DNA-binding domain protein	K	19	18	3	1
Putative cold-shock DNA-binding domain protein	K	7	11	9	1, 2, 3
NmrA family protein	G,M	56	134	78	
Dihydroxy-acid dehydratase	E,G	25	43	9	1
Pyruvate kinase	G	46	93	53	3
Acetyl-CoA acetyltransferase	I	72	126	86	2, 3
Pyruvate carboxylase	C	58	105	86	2
Pyridoxine biosynthesis protein	H	100	163	116	2, 3
Chaperone DnaJ domain protein	O	16	19	5	1
Chaperone protein DnaK	O	318	456	257	
Chaperonin GroEL	O	384	736	456	
Chaperonin GroEL	O	371	506	311	1
Chaperonin GroEL	O	243	312	195	1
SufBD protein	O	23	34	13	1, 3
Asp/Glu racemase	E	0	0	7	1
Porphobilinogen deaminase	H	28	20	14	1, 2
Porphobilinogen synthase	H	24	20	12	1, 2
Glycerol kinase	C	135	101	248	1
Glycerol kinase	C	62	50	88	1
Glycerone kinase	G	37	55	89	1
Isocitrate lyase	C	33	31	74	1
Malate synthase A	C	13	20	32	1
Malate synthase A	C	13	16	27	1
Glyoxylate reductase	C, H, R	3	1	32	1
Isocitrate dehydrogenase (NADP(+))	C	60	67	40	1, 2
Extracellular solute-binding protein, family 5	E	0	0	20	1
Extracellular solute-binding protein, family 5	E	0	0	4	1
Extracellular solute-binding protein, family 3	E, T	0	0	10	1
ABC transporter related	P	1	0	8	1
Putative ABC-type sulfonate transport system	P	0	0	11	1
Extracellular solute-binding protein family 1	P	21	41	6	1
Periplasmic binding protein	P	0	16	0	3
Periplasmic binding protein	P	1	16	0	3
Extracellular solute-binding protein, family 1	P	16	41	15	3
Conserved unknown protein	R	183	50	58	2
Unknown protein	none	31	35	76	1
Protein of unknown function DUF1684	S	5	4	37	1
Unknown protein	none	32	35	64	1
Unknown protein	P	16	24	37	1
Conserved unknown protein	none	19	18	39	1

I.D. Insignificant differences between the following growth states: 1) 4-CP and 4-NP, 2) 4-NP and phenol, 3) 4-CP and phenol.

[a]Genes in previously reported 4-chlorophenol degradation cluster[19].

The color legend is the same as in Table 3.

Table 6. Differences in protein expression (spectral counts) between wild type and mutant strains, both growing on phenol.

Protein description	COG	Wt	T99	I.D.
Monooxygenase (*cphC-I*)[a]	Q	470	529	
Maleylacetate reductase (*cphF-II*)[a]	C	196	139	
Hydroxyquinol 1,2 dioxygenase (*cphA-II*)[a]	Q	184	104	
Monooxygenase FAD-binding (*cphC-II*)[a]	C,H	161	108	
Hydroxyquinol 1,2 dioxygenase (*cphA-I*)[a]	Q	158	3	
carbohydrate kinase FGGY	C	88	118	
Isocitrate lyase	C	74	540	
flavin reductase domain protein FMN-binding (cphB)[a]	R	64	5	
Putative monooxygenase	C	62	0	
Conserved unknown protein	R	58	303	
Putative monooxygenase	C	57	0	
Unknown protein	M	47	5	
Protocatechuate 3,4-dioxygenase, alpha subunit	Q	44	17	
Taurine dioxygenase	Q	44	1	
Isocitrate dehydrogenase (NADP(+))	C	40	63	
Protein of unknown function DUF1684	S	37	0	
Aldehyde dehydrogenase	C	35	652	
Putative monooxygenase	C	33	0	
Glyoxalase/bleomycin resistance protein/dioxygenase	R	33	12	
Malate synthase A	C	32	42	*
Glyoxalase/bleomycin resistance protein/dioxygenase	S	29	8	
Malate synthase A	C	27	149	
Putative monooxygenase	none	27	0	
Extracellular solute-binding protein family 5	E	20	0	
Nitrilotriacetate monooxygenase component A	C	19	0	
Maleylacetate reductase (*cphF-I*)[a]	C	18	19	*
Putative monooxygenase	none	15	0	
Putative ABC-type sulfonate transport system	P	11	0	
Phenol hydroxylase domain protein dimerisation	C,H	11	5	*
Glyoxalase/bleomycin resistance protein/dioxygenase	none	11	4	
Extracellular solute-binding protein family 3	E,T	10	0	
ABC transporter related	P	8	0	
Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))	C	7	46	
Transcriptional activator domain (*cphS+cphR*)[a]	K	3	4	*
Probable molybdate ABC transporter (*cphX*)[a,b]		3	0	
Extracellular solute-binding protein family 1	G	2	90	
Catechol 2,3 dioxygenase	R	2	1	*
Conserved unknown protein	none	0	22	
ABC transporter related	V	0	12	

[a]Genes in previously reported 4-chlorophenol degradation cluster[19].

I.D. = Insignificant difference; * no significant difference between samples.

[b]Protein with no corresponding gene annotation in the draft genome.

The color legend and COG definitions are the same as in Table 3.

Table 7. Altered KEGG pathways in comparisons of wild type (wt) vs. mutant strains.

KEGG pathway	wt Higher	Mutant Higher
Citrate cycle	6.4.1.1	2.3.3.1
	1.3.99.1	4.2.1.3
	6.2.1.4	1.1.1.42
	6.2.1.5	1.2.4.2
		1.8.1.4
		2.3.1.61
Pyruvate metabolism	6.4.1.1	1.1.2.3
	2.3.1.9	1.2.2.2
		1.2.1.3
		6.2.1.1
		1.8.1.4
		1.1.1.40
		2.3.3.9
		1.1.99.16
Glyoxylate and dicarboxylate metabolism	1.2.1.2	2.3.3.1
	3.5.4.9	4.2.1.3
		4.1.3.1
		4.1.1.47
		2.3.3.9
		5.3.1.22
		2.7.1.31
Reductive carboxylate cycle (CO_2 fixation)	1.3.99.1	6.2.1.1
	6.2.1.5	4.2.1.3
		1.1.1.42
Purine metabolism	3.5.4.4	2.7.7.6
	2.4.2.1	1.17.4.1
	2.4.2.4	2.7.7.4
	3.5.1.5	
	1.7.1.7	
Pyrimidine metabolism	2.4.2.1	2.7.7.6
	2.4.2.4	1.17.4.1
		1.8.1.9
		6.3.4.2
Valine, leucine and isoleucine biosynthesis		2.2.1.6
		1.1.1.86
		4.2.1.9
		6.1.1.9
Pantothenate and CoA biosynthesis	2.7.8.7	2.2.1.6
		1.1.1.86
		4.2.1.9
1,4-Dichlorobenzene degradation	1.14.13.20	
	1.3.1.32	
	1.13.11.37	
Benzoate degradation via CoA ligation	1.1.1.35	
	1.3.99.1	
	3.5.1.4	
	2.3.1.9	
	4.2.1.17	
Benzoate degradation via hydroxylation	3.1.1.24	
	2.8.3.6	
	1.3.1.32	
	1.13.11.37	
	1.13.11.3	

EC-numbers of proteins that are upregulated in the wild type strain compared to the mutant (left column); upregulated in the mutant strain compared to the wild type (right column).

Figure legends

Figure 1. Heat map showing differences in protein expression. a) Growth conditions: Upper tier: Green = Phenol, Red = 4-NP, Black = 4-CP; Middle tier: Black = 5 C, Red = 28 C; Lower tier: Black = Wild type strain, Red = Mutant strain. b) 1678 genes showing significantly different protein abundances in pairwise comparisons between treatments according to spectral count measurements. Individual genes are represented by a single row, and each replicate culture for each growth condition by a single column. Each cell represents the expression level of a protein under one growth condition, relative to the mean expression level across all conditions. Red represents over-expression, and green represents under-expression. Black cells were not significantly different. c. Tree diagram of cluster analysis showing similarities between samples, using the same color scheme corresponding to the columns given in (a).

Figure 2. 4-chlorophenol degradation pathway, possibly also used for degradation of 4-NP[3], and the corresponding *cph* gene cluster in *A. chlorophenolicus*. An arrowhead indicates the cphA-I gene that was disrupted by transposon mutagenesis in the mutant strain and the corresponding protein in the pathway is crossed out. Adapted from[19].

Figure 3. Comparisons of glyoxylate and dicarboxylate metabolism for cultures grown on phenol compared to 4-NP or 4-CP (a) and the wild type and mutant strains (b). a) EC-numbers marked green indicate proteins that are up-regulated in phenol-grown cultures compared to cultures grown on 4-NP or 4-CP. b) EC-numbers marked green indicate proteins that are up-regulated and red EC-numbers proteins that are down-regulated in the wild type strain compared to the mutant when both are grown on phenol. Grey EC-numbers are those without corresponding proteins in ENZYME database.

Figure 1

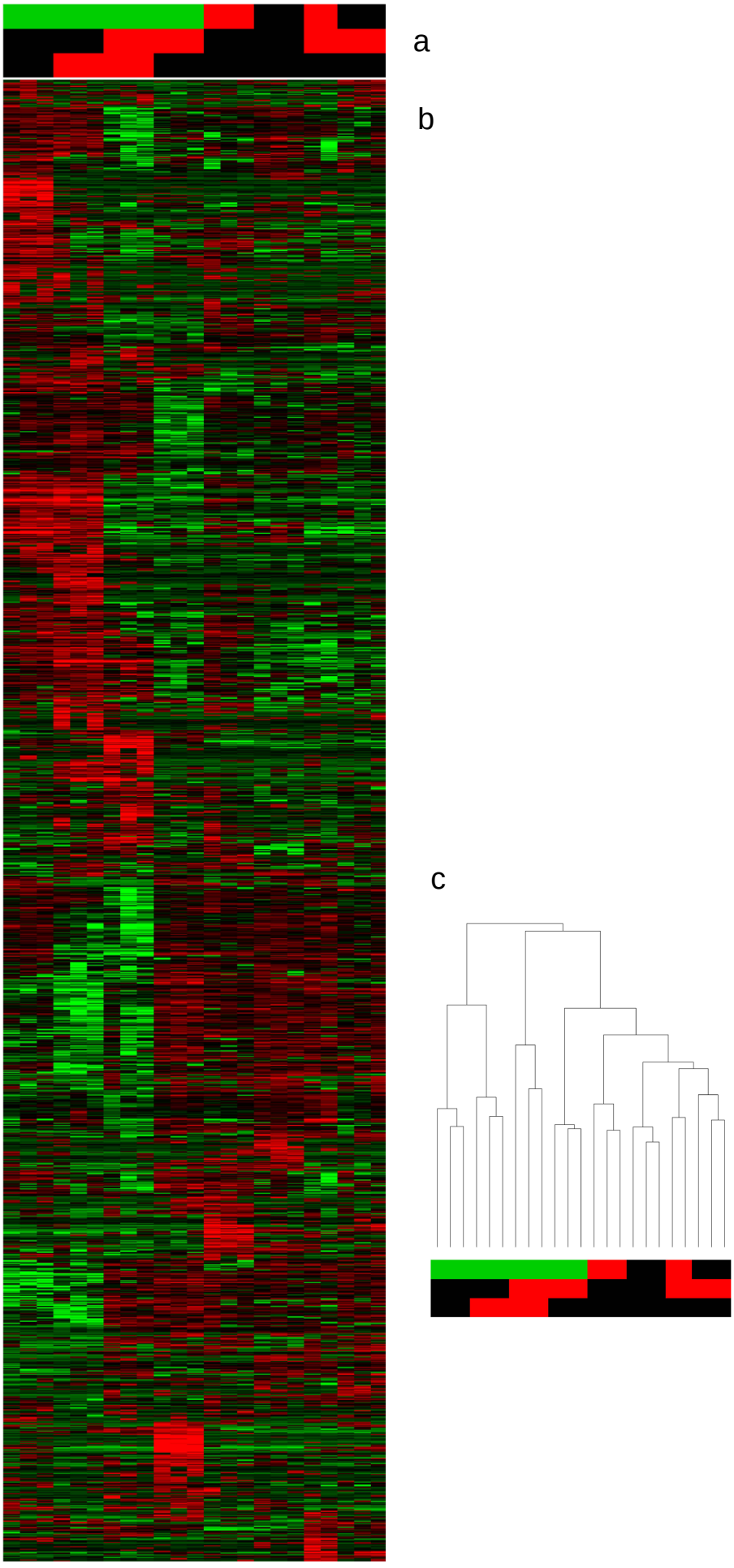


Figure 2

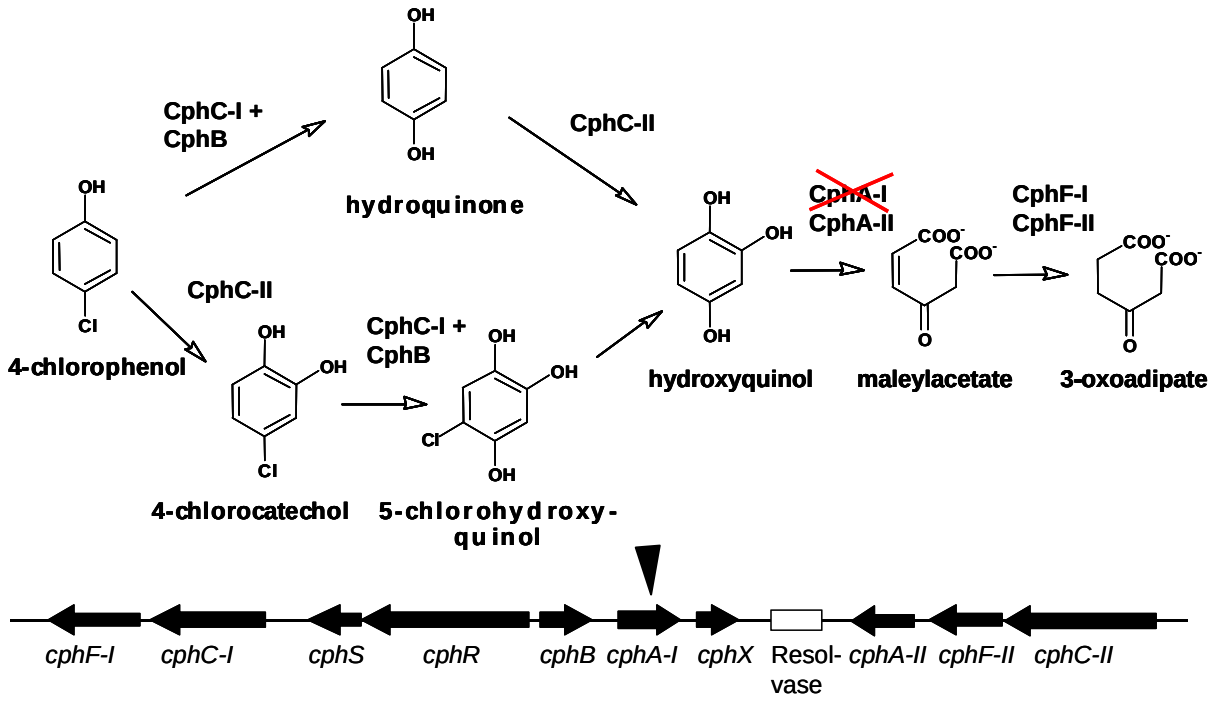
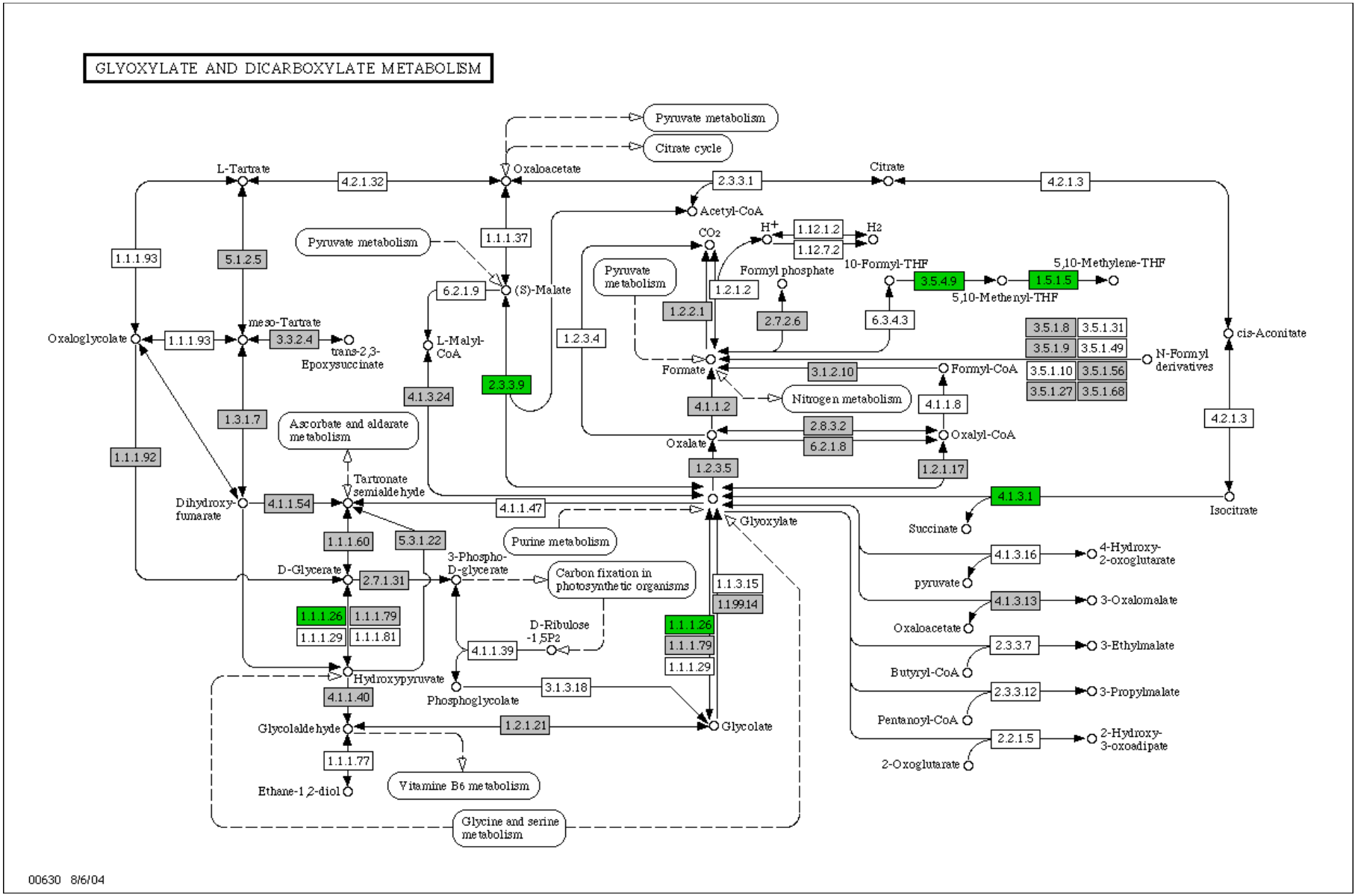
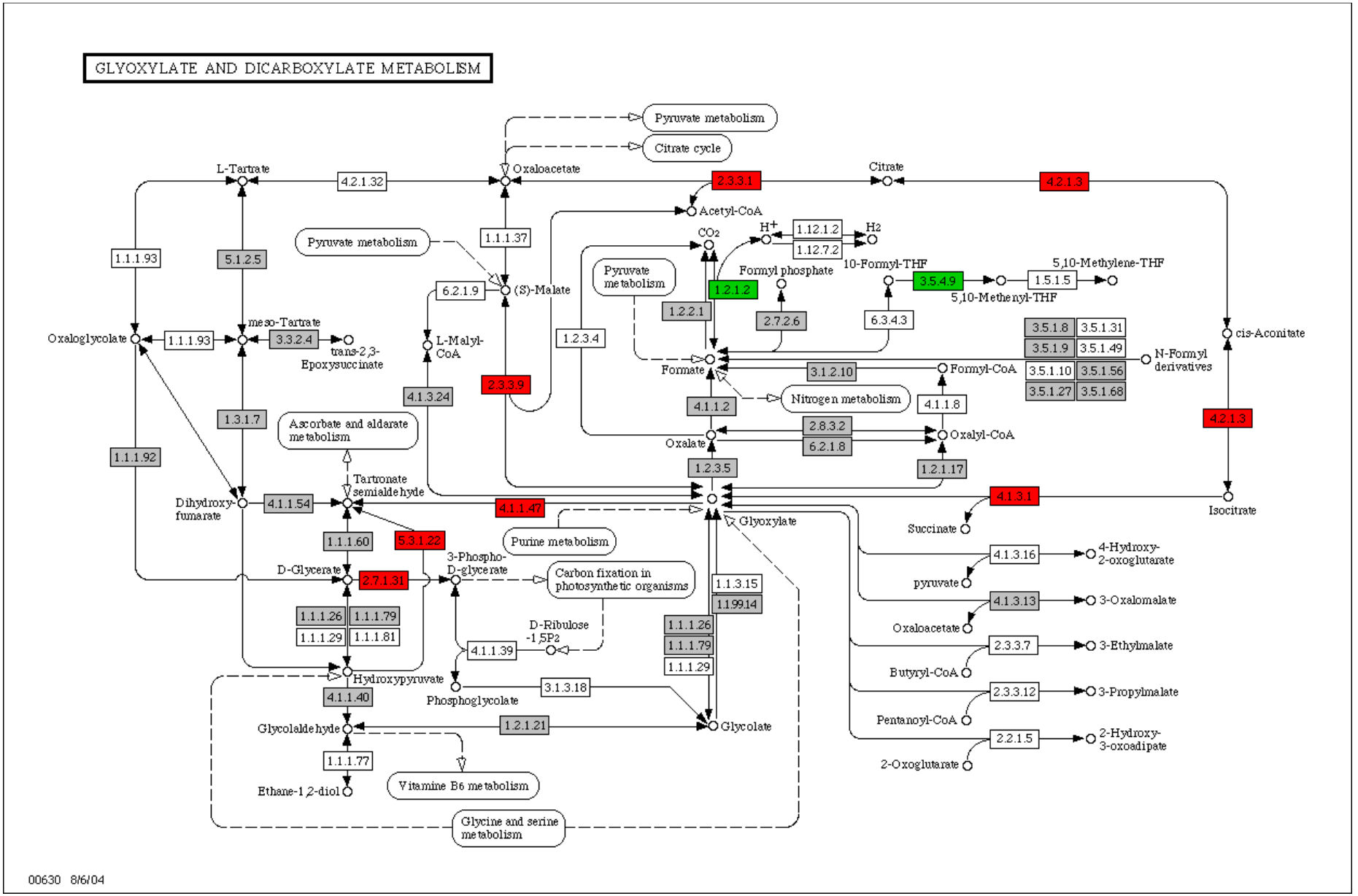

CphC-I +
CphB
hydroquinone
CphC-II
4-chlorophenol
CphC-II
CphC-I +
CphB
4-chlorocatechol
5-chlorohydroxy-
quinol
hydroxyquinol
CphA-I
CphA-II
maleylacetate
CphF-I
CphF-II
3-oxoadipate
cphF-I
cphC-I
cphS
cphR
cphB
cphA-I
cphX
Resol-
vase
cphA-II
cphF-II
cphC-II

Fig3

a


GLYOXYLATE AND DICARBOXYLATE METABOLISM
Pyruvate metabolism
Citrate cycle
L-Tartrate
4.2.1.32
Oxaloacetate
2.3.3.1
Citrate
4.2.1.3
Acetyl-CoA
1.1.1.93
5.1.2.5
Pyruvate metabolism
1.1.1.37
CO2
H+
1.12.1.2
1.12.7.2
H2
10-Formyl-THF
5,10-Methylene-THF
3.5.4.9
1.5.1.5
5,10-Methenyl-THF
Formyl phosphate
6.2.1.9
(S)-Malate
Pyruvate metabolism
1.2.1.2
1.2.2.1
2.7.2.6
6.3.4.3
Oxaloglycolate
1.1.1.93
meso-Tartrate
3.3.2.4
trans-2,3-Epoxysuccinate
L-Malyl-CoA
1.2.3.4
3.5.1.8
3.5.1.31
3.5.1.9
3.5.1.49
N-Formyl derivatives
cis-Aconitate
Formate
3.1.2.10
Formyl-CoA
3.5.1.10
3.5.1.56
3.5.1.27
3.5.1.68
4.1.3.24
2.3.3.9
Nitrogen metabolism
4.1.1.8
4.1.1.2
1.3.1.7
Ascorbate and aldarate metabolism
2.8.3.2
6.2.1.8
Oxalate
Oxalyl-CoA
4.2.1.3
1.1.1.92
1.2.3.5
1.2.1.17
Tartronate semialdehyde
Dihydroxy-fumarate
4.1.1.54
4.1.1.47
4.1.3.1
Isocitrate
Glyoxylate
Succinate
Purine metabolism
1.1.1.60
5.3.1.22
3-Phospho-D-glycerate
4.1.3.16
4-Hydroxy-2-oxoglutarate
D-Glycerate
2.7.1.31
Carbon fixation in photosynthetic organisms
pyruvate
1.1.3.15
1.1.99.14
4.1.3.13
3-Oxalomalate
1.1.1.26
1.1.1.79
1.1.1.29
1.1.1.81
D-Ribulose-1,5P2
Oxaloacetate
1.1.1.26
1.1.1.79
1.1.1.29
2.3.3.7
3-Ethylmalate
4.1.1.39
Butyryl-CoA
Hydroxypyruvate
Phosphoglycolate
3.1.3.18
2.3.3.12
3-Propylmalate
4.1.1.40
Pentanoyl-CoA
Glycolaldehyde
1.2.1.21
Glycolate
2.2.1.5
2-Hydroxy-3-oxoadipate
1.1.1.77
2-Oxoglutarate
Ethane-1,2-diol
Vitamine B6 metabolism
Glycine and serine metabolism
00630 8/6/04

b


GLYOXYLATE AND DICARBOXYLATE METABOLISM
Pyruvate metabolism
Citrate cycle
L-Tartrate
4.2.1.32
Oxaloacetate
2.3.3.1
Citrate
4.2.1.3
Acetyl-CoA
1.1.1.93
5.1.2.5
Pyruvate metabolism
1.1.1.37
CO2
H+
1.12.1.2
1.12.7.2
H2
Pyruvate metabolism
Formyl phosphate
10-Formyl-THF
5,10-Methylene-THF
6.2.1.9
(S)-Malate
1.2.1.2
3.5.4.9
1.5.1.5
5,10-Methenyl-THF
1.2.2.1
2.7.2.6
6.3.4.3
Oxaloglycolate
1.1.1.93
meso-Tartrate
3.3.2.4
trans-2,3-Epoxysuccinate
L-Malyl-CoA
1.2.3.4
3.5.1.8
3.5.1.31
3.5.1.9
3.5.1.49
N-Formyl derivatives
cis-Aconitate
Formate
3.1.2.10
Formyl-CoA
3.5.1.10
3.5.1.56
3.5.1.27
3.5.1.68
2.3.3.9
4.1.3.24
Nitrogen metabolism
4.1.1.2
4.1.1.8
1.3.1.7
Ascorbate and aldarate metabolism
2.8.3.2
6.2.1.8
Oxalate
Oxalyl-CoA
4.2.1.3
1.1.1.92
Tartronate semialdehyde
1.2.3.5
1.2.1.17
Dihydroxy-fumarate
4.1.1.54
4.1.1.47
Glyoxylate
4.1.3.1
Isocitrate
Succinate
1.1.1.60
5.3.1.22
Purine metabolism
4.1.3.16
4-Hydroxy-2-oxoglutarate
D-Glycerate
2.7.1.31
3-Phospho-D-glycerate
Carbon fixation in photosynthetic organisms
1.1.3.15
1.1.99.14
pyruvate
4.1.3.13
3-Oxalomalate
1.1.1.26
1.1.1.79
1.1.1.29
1.1.1.81
D-Ribulose-1,5P2
4.1.1.39
1.1.1.26
1.1.1.79
1.1.1.29
Oxaloacetate
2.3.3.7
3-Ethylmalate
Hydroxypyruvate
Butyryl-CoA
2.3.3.12
3-Propylmalate
4.1.1.40
3.1.3.18
Phosphoglycolate
Pentanoyl-CoA
Glycolaldehyde
1.2.1.21
Glycolate
2.2.1.5
2-Hydroxy-3-oxoadipate
1.1.1.77
2-Oxoglutarate
Ethane-1,2-diol
Vitamine B6 metabolism
Glycine and serine metabolism
00630 8/6/04