Discussion
Discussion

Over expression and knockout of genes are performed to establish new metabolic pathways, to increase the metabolic flux towards the desired product, decrease or shut down the undesirable metabolic flux, and to balance the metabolic flux which all contribute to enhance the metabolic fluxes towards optimal and desired product formation. One of the open questions in the quest for genetic targets is how multiple targets are identified. In principle, there should be a set of genes, appropriately expressed which would help in conferring the optimal phenotype. However, such an optimal set cannot be experimentally discovered as it is impossible to construct and evaluate all the possible combinations of these genes and their expression levels. Therefore, a proper search is required to identify multiple gene targets (Stephanopoulos, et al., 2007).

There are several approaches for the modifications of the E. coli host to achieve the desired goal (Chou, 2007). Since it is a fairly easy task to create mutant expression libraries in E. coli, the more critical aspect is the design of screening protocols to select for desirable mutants. The most common technique has been to screen for 'fitter' organisms like those with higher growth rate. These selections can be done in different environments for e.g. in a CSTR in the presence of different toxic compounds. The cells which are able to outgrow the competition in the presence of a toxin get selected.

One example of this is the isolation of mutant strains of Saccharomyces cerevisiae that are resistant to high acetate concentrations by Gilbert et al. (2009). Acetate is present at high concentrations, over 10 g/L acetate, in lignocellulosic hydrolysates (Taherzadeh et al., 1997; Klinke et al., 2003) and inhibits the growth of S. cerevisiae (Pampulha and Loureiro-Davis, 1989; Palmqvist et al., 1999). The cytostat cultivation technique enabled isolation of acetate tolerant strains of S. cerevisiae in less than 5 days. The mutant strains produce ethanol at higher titers and higher rates than both the parental strain and a commercially available ethanologenic strain. Analysis by cDNA microarrays revealed differential gene amplifications in each isolated mutant.

This evolutionary method of selection has many advantages such as the cytostat cultures grow at low cell densities and the maximum specific growth rate supported by the feed medium, leading to larger absolute growth advantages for the mutants over the parental strains. The result is fewer generations are required, and the generations occur on a shorter time scale. Even in cases where the genetic modifications for tolerance are known, cytostat technology, resulting in mutant isolation in only a few days, is extremely competitive with the time requirements of standard molecular biology techniques. Cytostat technology also has the advantage of utilizing the native
Discussion

genome of the organism, such that phenotypic improvements are likely to be stable with comparison to improvements conferred by recombinant nonnative genes. The genetic basis of phenotypes have traditionally been identified by genetic selections and/or screens followed by Sanger sequencing. Approaches include random mutagenesis and selection or high throughput screening. In these experiments the ability to probe every gene present in the organism is limited by the number of cells that can be screened and the number of targets that can be sequenced. Therefore, in addition to being labor intensive and costly, these traditional methods will probably not identify many relevant genes, thus producing an incomplete picture of the genetics underlying a complex phenotype. Furthermore, if a few relevant genes are identified, evaluating their impact in totality requires creating a combinatorial library of all genetic combinations for further analysis, which makes the problem unmanageable as the number of relevant genes increase.

Other methods involve metabolic engineering but considering the general feasibility of introducing any heterologous genes (natural and synthetic) and of making any change in the host genome, the set of genetic possibilities available to the metabolic engineer is also infinite. Only a small number of these infinite possibilities will be effective in achieving the metabolic engineer's goal. This also means that metabolic engineering is almost certain to fail unless powerful algorithms can be identified which greatly increase above random chance the probability of identifying an effective genetic change. The classical problem in the early emergence of metabolic engineering is identifying a flux-limiting step in a specified metabolic pathway. This formulation of the objective embodies an implicit assumption of several layers of knowledge about the pathway. Not only is the identity of the pathway assumed, but also the identity of the catalysts involved should be known. Furthermore, to choose a possible flux-limiting step, other than one at random, much more information must be available, whether in terms of reaction kinetics, intermediate metabolite concentrations, or results from well-designed stimulus response experiments (Cornish-Bowden and Cardenas, 1990; Galazzo and Bailey, 1990; Schlosser et al., 1993).

However, the bigger challenge is to design overproducers of the product of interest rather than cells which are more 'fit' than their counterparts since the 'fit' cells only channel metabolic fluxes more efficiently to biomass formation. We therefore decided to reverse the standard strategy and look for 'less fit' or non-growing cells. The major problem with this method is that there is no simple enrichment strategy which would reduce the number of samples to be studied. We had to individually screen single colonies on master plates to get the desired phenotype of slow growers and later
check it in shake flasks with different media compositions. More importantly not all 'slow-growers' would be interesting/ or provide useful leads. The most promising candidates would be those where slow growth and high product formation would be observed. The likelihood of getting such useful mutants was low, even then we did get a few promising candidates. More importantly many knock downs tended to hamper both growth and product formation. These knock-downs would be useful if they also turned out to be the rate-limiting step in product formation. That is why the transcriptomic analysis was done in tandem with this work to observe whether the genes identified by inverse metabolic engineering were also downregulated during induction. But we observed that the transcriptomic profiles of these 21 candidates identified did not give us any clear idea of their regulations. All the genes had a similar pattern of getting downregulated in the 4th hour and then recovering in the 6th hour. So these leads (which were generated by the inverse metabolic engineering approach) could not have been generated by the known theoretical metabolic models or even from the transcriptomic or proteomic analysis. Thus, we have been able to design a strategy which helps us select better producers and lead us to a gene identification which was both non-obvious and counter-intuitive.

An interesting refinement in our study was that the genomic library was prepared using two different vectors (with different characteristics) by ligating 200-500 bp fragments of digested *E. coli* genomic DNA (which would work in the antisense fashion by binding to the corresponding part of the complementary region of its gene and block its translation) in high copy number plasmid pRSET A with a very strong ‘T7’ promoter and low copy number plasmid pBAD33 with a comparatively weak promoter ‘ara’. The weak promoter was used so that the antisense RNA’s produced does not create a metabolic burden to the cells but has its effect in moderation. Those transcripts identified from the screening of the pRSET A library and showing an increase in recombinant protein production upon induction, should be 'knocked-out' from the *E. coli* genome but those transcripts identified from the pBAD33 library needed to be only ‘knocked-down’ since the pBAD33 carried a titratable promoter. We hypothesized that some of the clones would carry partial copies of the *E. coli* genes in the reverse orientation and hence produce antisense RNA which upon expression would down regulate the corresponding gene and simultaneously tend to partially down regulate some pathways.

After thorough screening of the libraries we identified 21 clones (Table 4) which were responsible for negligible growth but had sufficient metabolic activity and in some cases enhanced GFP production. There is thus a distinct possibility that these clones
would have sustained metabolic activity in a non-growing quiescent mode which would increase the operational life-span of the bio-processes and improve process economics by decoupling production from cell growth as has been reported by other researchers (Sonderegger, 2005).

GN11 was selected for the further studies containing an insert of 403 bp of the ribB gene (3, 4 dihydroxy-2-butanolone-4-phosphate synthase) in the reverse direction. It is synthesized from ribulose-5-phosphate and serves as the biosynthetic precursor for the xylene ring of riboflavin. It is bifunctional enzyme with GTP cyclohydrase 11 that catalyzes the first committed step in the biosynthesis of riboflavin. Not much is reported about the E. coli ribB gene however, in B. subtilis the genes encoding the riboflavin biosynthetic enzymes were found to be clustered in a single 4.3 kbp operon (rib operon) (Mironov et al., 1989; Mironov et al., 1990; Perkins and Pero, 1993). The gene products of the rib operon (RibG, RibB, RibA and RibH) catalyze the conversion of GTP and ribulose-5-phosphate to riboflavin (Bacher, 1991; Bacher et al., 1993). It is also reported that ribC mutant B. subtilis overproduces riboflavin. However the ribB mutant in E. coli leads to slow growth but remains metabolically active (Raina et al., 1991), which corroborates with our screening studies. This suggests that by blocking ribB gene, possibly some different metabolic pathways were blocked which was not essential for the cell survival and the metabolic flux was diverted towards recombinant protein production. This hypothesis is supported by the shake flask studies in which when there was knockdown of this gene, specific yield of GFP was 347 AU/gm DCW in contrast to that of control, where it was 47 AU/gm i.e. an increase of 7 fold was achieved.

Despite advances in mixing and air dispersion technology, in membrane systems, and in discovery of various additives for media which can enhance oxygen transfer, limitation of desired cellular activities by oxygen supply remains an important concern in many situations. Another crucial problem faced during high cell density cultivation of E. coli cells is the production of acetic-acid as an extracellular co-product of aerobic fermentation, and this exists as the ion acetate at the neutral pH used in E. coli fermentations (Eiteman et al., 2006). Both the problems have been addressed many a times in the past using various approaches but we tried solving them simultaneously so that a single chromosomal integration event would serve to both knock-out and knock-in desirable genes.

Thus the metabolic changes done previously can be complemented by introducing foreign genes from other hosts and enhancing the metabolic pathways further. It has been shown that in HCDC, cells get oxygen limited inspite of maintaining high D.O.
levels in the bioreactor. Micro mixing problems lead to the creation of microaerophilic or oxygen starvation conditions. Thus the genes for oxygen uptake are often upregulated and one such stride by us was introduction of the gene \textit{vhb}. The co-expression of the \textit{Vitreoscilla} gene encoding bacterial hemoglobin, VHb, has been used to enhance foreign protein formation under microaerobic conditions in which cellular $O_2$ levels may be limited (Kalio and Bailey 1988; Webster \textit{et al.}, 1988; Enayati \textit{et al.}, 1990; Kang \textit{et al.}, 2002). Hence, the introduction of the vhb gene in the host was to shun the deficiency of oxygen and the inability of the host cells to take up oxygen during overproduction of recombinant proteins in high cell density culture.

In this study, we tried to avoid the acetate accumulation problem by knocking out the acetate operon (encoding the genes \textit{ackA, pta} and \textit{yfcC}) from the \textit{E. coli} genome while keeping its promoter intact. Under this intact promoter we simultaneously knocked-in the \textit{vhb} gene. The acetate promoter's transcriptional activation is regulated by FNR-ArcA dual regulator. FNR stimulates the transcription of many genes required for fermentation and anaerobic respiration and also represses the transcription of some genes that function only during aerobic growth. This system is active only during anaerobic growth and its activity is regulated by oxygen. On the other hand, ArcA regulator represses the gene expression of respiratory enzymes and induces the expression of proteins that allows the activity of central metabolism enzymes sensitive to oxygen loss, as well as that of some enzymes involved in fermentative metabolism and works under microaerobic conditions. Thus Vhb would be produced when this promoter gets activated automatically under oxygen deficiency, and the modified strain was named CG1. In oxygen sufficient conditions in the shake flasks, the fluorescence of the control strain was 1.26 fold higher than the strain CG1 and specific yield in turn was 1.08 fold higher. Whereas, in the oxygen deficient conditions, GFP expression level of the strain CG1 was 1.3 fold higher and the specific yield of the protein was 1.5 fold higher than the control strain. Moreover, CG1 accumulated less acetate in high cell density culture because of the knock-out of the genes.

From the inverse metabolic engineering studies we had earlier got a lead that blockage of the \textit{ribB} gene could increase protein production. Further modifications were therefore made in the strain CG1 by knocking out this \textit{ribB} gene from its genome, resulting in a strain called CG2.

Shake flask studies showed CG1 was better than control in an oxygen deficient condition but CG2 was better in both oxygen sufficient or deficient conditions. The specific yield of CG2 was 1.7 fold higher than the unmodified strain and 1.1 fold higher than the strain CG1. After successfully achieving a fairly high specific yield in shake
flasks by the modified strains CG1 and CG2 compared to the control strain, the next logical step was to translate this into production in a bioreactor. The productivity of the strains CG1, CG2 and control were tested in bioreactors run in batch and fedbatch mode.

In the batch reactor, both the volumetric product concentration and specific product yield of CG2 increased from the control. \( q_p \) could also be sustained for a longer period of time in the CG2 strain. Moreover, the fall in rpm was much less in the modified strain thus proving the better oxygen utilization by the cells. Due to the presence of the Vhb protein and the removal of the acetate genes, the strain could perform better by sustaining more stress. The blockage of the \( \text{ribB} \) gene also led to the diversion of flux towards recombinant protein production.

For the fed batch run, higher specific growth rates also lead to oxygen starvation even at low cell densities as well as acetate accumulation. Thus the advantages of high \( \mu \) are lost and often an optimum \( \mu \) is postulated for recombinant protein production which is high enough to give good cellular health but not too high to promote oxygen starvation or acetate formation. However, in our case we had introduced \( vhb \) gene and simultaneously blocked the acetate production pathway therefore, it was hypothesized that this strain could be a better grower even at a much higher \( \mu \). Thus fedbatch was run at two different specific growth rates by controlling the rate of feed. In both cases oxygen was not splurged so as to create an oxygen deficient condition in the high cell density cultures.

In the fed batch where specific growth rate was 0.6 h\(^{-1}\), glycerol in the feed was replaced by glucose. When excess glucose is supplied, carbon flux through glycolysis exceeds the capacity of TCA cycle. Therefore, by-products (e.g. acetate) are secreted under both anaerobic and aerobic conditions (Majewski & Domach 1990). These accumulated by-products can inhibit cell growth and recombinant protein production (Holms 1986, Hahm et al., 1994). Also, high glucose concentrations can decrease production of foreign protein through induction of catabolite repression on some types of transcriptional promoters. So a highly stressed condition was created, such that our modified strain could prove itself. The major success of the fed-batch experiments can be gauged by the \( q_p \) vs. time profile, post induction. It is clear that there was very little change in the pattern as well as the absolute \( q_p \) values of the strain CG1 and CG2 but was 3 fold higher from the control strain. That is possibly because cells were healthier in the modified strain and the rate of formation of product was much higher than the growth rate whereas in the control, \( q_p \) fell to zero when \( \mu \) was also low, possibly
Discussion

because once cellular health was poor and recombinant protein expression was shut off.

Understanding the effects of recombinant protein overexpression on global genomic regulation is important for designing metabolic engineering strategies. But most-of the microarray based studies have been carried out under not clearly defined conditions in flask cultures, where overlaps of different responses may have occurred (Richmond et al., 1999; Arnold et al., 2001). Nor are there any reports of comparing the transcriptomic data across two different μ's. The best defined homogeneous cell material could be gained from tightly controlled chemostat cultivation. To accomplish this objective the transcriptome response of recombinant cultures of the modified strain under two different specific growth rates and at different time points were analyzed. The transcriptome levels were determined using Affymetrix E. coli Antisense DNA microarrays, such that the entire genome was evaluated. These two transcriptome responses were also compared to recombinant culture of the control strain. Moreover, we have in our lab the transcriptomic profiles of a protein IFN β which forms inclusion bodies and xylanase which is a soluble protein. Thus a comparison of soluble protein with an inclusion body forming protein across different range of specific growth rates can be compared further which has not been done earlier, to identify various steps where metabolic engineering should be targeted. Some of the interesting findings of the transcriptomic analysis were the completely different time profile of the key genes involved in TCA cycle, oxidative stress, amino acid and nucleotide biosynthesis and energy metabolism pathways. Clearly the modified strain had a distinctly different stress response possibly because of growth stoppage. This strain needs to be tested for the product of other recombinant proteins. Future systemic analysis would provide us leads to the genetic manipulations for a better host platform for the overexpression of recombinant protein.