

Genome mining, in silico validation and phase selection of a novel aldo-keto reductase from *Candida glabrata* for biotransformation

Souvik Basak, Nanda Gopal Sahoo & Angayar K Pavanagam

To cite this article: Souvik Basak, Nanda Gopal Sahoo & Angayar K Pavanagam (2017): Genome mining, in silico validation and phase selection of a novel aldo-keto reductase from *Candida glabrata* for biotransformation, *Bioengineered*, DOI: [10.1080/21655979.2017.1342911](https://doi.org/10.1080/21655979.2017.1342911)

To link to this article: <http://dx.doi.org/10.1080/21655979.2017.1342911>



Accepted author version posted online: 23 Jun 2017.



Submit your article to this journal [↗](#)



Article views: 9



View related articles [↗](#)



View Crossmark data [↗](#)

Genome mining, *in silico* validation and phase selection of a novel aldo-keto reductase from *Candida glabrata* for biotransformation

Souvik Basak^{a,*}, Nanda Gopal Sahoo^{b,*}, Angayar K Pavanam^c

^aDr. B.C. Roy College of Pharmacy & Allied Health Sciences, Durgapur, WB, India

^bNanoscience and Nanotechnology Centre, Department of Chemistry, Kumaun University, Nainital, Uttarakhand, India

^cInternational College of Engineering and Management (University of Central Lancashire), Muscat, Oman

***Authors for correspondence:** Souvik Basak; Email: souvik_basak1@yahoo.com

***Authors for correspondence:** Nanda Gopal Sahoo; Email: ngsahoo@yahoo.co.in

Previously, we published cloning, over-expression, characterization and subsequent exploitation of a carbonyl reductase (*cr*) gene, belonging to general family aldo-keto reductase from *Candida glabrata* CBS138 to convert keto ester (COBE) to a chiral alcohol (ethyl-4-chloro-3-hydroxybutanoate or CHBE). Exploiting global transcription factor CRP, rDNA and transporter engineering, we have improved batch production of CHBE by trinomial bioengineering. Herein, we present the exploration of *cr* gene in *Candida glabrata* CBS138 through genome mining approach, *in silico* validation of its activity and selection of its biocatalytic phase. For exploration of the gene under investigation, three template genes were chosen namely *Saccharomyces cerevisiae* YDR541c, YGL157w and YOL151w. The CR showed significant homology match, overlapping of substrate binding site and NADPH binding site with the template proteins. The binding affinity of COBE towards CR (-4.6 Kcal/ mol) was found higher than that of the template proteins (-3.5 to -4.5 Kcal/ mol). Biphasic biocatalysis with cofactor regeneration improved product titer 4~5 times better than monophasic biotransformation. Currently we are working on DNA Shuffling as a next level of strain engineering and we demonstrate this approach herein as a future strategy of biochemical engineering.

Keywords

cofactor regeneration, homology match, docking, binding site, DNA shuffling

Introduction

In previous studies, many enzymes of Aldo-keto Reductase (AKR) and Carbonyl reductase (CR) family were cloned, characterized and used in the asymmetric synthesis of (*S*)-CHBE¹⁻⁴ and (*R*) CHBE⁵. AKR & CR find extensive applications in pharmaceutical industry. For example, they are employed as key chiral intermediates in the enantioselective synthesis of slegenins B and C, they serve as two potential compounds against murine leukemia, they are also used for synthesis of HMG-CoA reductase inhibitors (hypolipidemic agents) and can be converted into 1,4 dihydropyridine type blocker (antihypertensive agents)¹⁻⁵. Similarly, they are also used for conversion of ethyl-4-chloro-3-oxobutanoate (COBE) to optically active ethyl-4-chloro-3-hydroxybutanoate (CHBE) as CHBE serve as a versatile precursor for pharmalogically valuable products¹⁻⁵. Although attempts have been made to augment bioconversion by either genetically manipulating the biocatalytic system with cofactor regeneration or fabricating the reaction media with single or multiple solvents, productivity has often faced shortfall due to obtaining higher reaction rate only at small substrate concentration (5~230 mM) thus eliciting a limiting batch output within the reactor⁶⁻⁹. Hence, establishment of a biocatalysis system with industrial competence has always been a prime search through the years which has driven researchers to find out newer proteins with improved activity and higher productivity.

In this context, in our previous publication¹⁰, we have reported the cloning, expression and purification of a carbonyl reductase from *Candida glabrata* CBS 138. We have also introduced the recombinant gene into globally engineered strain constructed by manipulating global transcription factor CRP. The improved tolerance by the host cell against organic phase, which had been added as an essential component of the biotransformation, led the output of the process shoot outstandingly higher. However, detailed methodology of the gene exploration together with optimization of the biocatalysis medium had not been described in details in the previous report. Thus, exploration of the gene through bioinformatics guided approach and its validation through docking studies has been presented in this work. In addition, this work also focuses on enhancing cell phenotype through DNA shuffling. Preliminary data obtained from DNA

shuffling appears to be promising with better product titer and hence it is envisaged to extend our research on improvement of microbial cell factory through the above mentioned techniques.

Discovery of CR protein from *Candida glabrata* CBS138

Three open reading frames namely *S. cerevisiae* YDR541c, YGL157w and YOL151w, reported for encoding aldo-keto reductases¹¹, were subjected to Protein BLAST (BLASTP) along with two crystallographically elucidated proteins of the same class namely aldehyde reductase 2 from *Sporobolomyces salmonicolor* AKU4429 (PDB ID: 1UJM) and an aldehyde reductase from *Sporidiobolus salmonicolor* (PDB ID: 1Y1P). The BLASTP results from all attempts revealed the repetitive hypothetical protein from *Candida glabrata* CBS138 (Protein ID: XP_445913.1). The BLAST scores revealed that the yeast proteins bear a plausible 55% to 62% identity to the target protein together with a staggering 98% sequence coverage with the latter. In contrast, 1UJM and 1Y1P possessed only 30% identity with the CR protein, however their 95% sequence swap with the target protein led us to a reliable approximation that XP_445913.1 from *Candida glabrata* CBS138 might belong to the same family as *S. cerevisiae* YDR541c, YGL157w, YOL151w together with 1UJM and 1Y1P. Multiple sequence alignment of the structures showed that they preserve a high degree of homology match with each other (**Fig. 1**) with a comprehensive amount of conserved amino acids at most of the positions.

Homology Modeling

For further investigations of XP_445913.1 from *C. glabrata* CBS138, homology model of the same was constructed using five template proteins as used for BLAST. In this process, primary homology models have been created for *S. cerevisiae* YDR541c, YGL157w and YOL151w since crystallographic structure has not been elucidated till date for these proteins. Homology modeling of all the four proteins were acquiesced using MODELLER9.12 (www.scilab.org/modeller). The overall folding of the homology model structure was same (RMSD 0.466) as analyzed by swiss pdb.

A further insight and analogical comparison of the modeled structure of XP_445913.1 with other proteins exhibited that the conserved catalytic residues such as S134, Y175 and K179 were similar to the

crystal structure of carbonyl reductase from *S. salmonicolor* and other yeast proteins (**Fig. 2**). In addition, similarities have been obtained for a lot of other amino acids too such as amino acids spanning hydrophobic channel of the two proteins, such as Phe 94→Phe 97, Trp 226→Val229, Pro241→Ala238, Ileu172→Leu174. However, for XP_445913.1, Q168 and E244 make the beginning of the channel little less hydrophobic than the crystal structure (1UJM) where there is Pro170 and Leu 241 in the equivalent positions (**Table 1**).

For estimation of modeling parameters, several parameters of our model has been performed such as ANOLEA (Atomic Non Local Environment)¹², QMEAN (Qualitative Model Energy Analysis)¹³, GROMOS (Groningen Molecular Simulation Computer Simulation Package)¹⁴. While ANOLEA calculates knowledge based distance dependent mean force potential, QMEAN evaluates the quality of the model based on certain scoring function and GROMOS is the force field based on molecular simulation. PROCHECK¹⁵ has been used for Ramachandran Plot. The modeled protein quality check with ANOLEA, QMEAN and GROMOS have been provided in **Fig. 3**. The model quality assessment has been performed using the Swiss-model workspace¹⁶.

Ramachandran plot has been performed in order to assess the model quality by analyzing the favored, allowed and generously allowed perturbations of residue-residue interaction (**Fig. 4**). From the plot, it can be observed that most of the residues have been clustered in α and β regions with a very few outliers suggesting that most of the interactions are favored folding interactions. Total quality of the model is assessed based on the reliability model and was found to be QMEAN6 of 0.528. The coloring residues plot with respect to errors (**Fig. 5A**) and normalized QMEAN6 plot with respect to query protein residues (**Fig. 5B**) have been provided to demonstrate the CR model accuracy. The pseudo energy plot of the contributing terms (**Fig. 5C**) has also been provided with their *z-scores* (with respect to the scores obtained from high-resolution structures in this protein subset). The scores obtained from high-resolution structures solved by X-ray crystallography has been taken as baseline score.

Docking studies

The docking studies were performed in AutoDockVina to evaluate the binding affinity of the substrate with the enzyme. The binding affinity of the target enzyme was compared to that of aforementioned standard AKR proteins to elucidate the enzyme potential as aldo-keto reductase against COBE. The energy minimized structure of COBE was prepared and converted to PDBQT format through MGL Tools 1.5.6. The modeled protein structures were freed from water molecules and inbound ligand by Discovery Studio 3.5 Visualizer. All the bonds and torsional angles of the ligand were allowed to rotate freely. *C. glabrata* CR was found to have more binding affinity (-4.6 Kcal/mol) compared to other same family of proteins (binding affinity ranging from -3.5 to -4.5 Kcal/mol) (**Table2**). A binding site analysis displayed that the substrate can fit nicely into the hydrophobic cavity of the enzyme and the amino acids especially T111 and Y175 can form hydrogen bonds with the substrate carbonyl oxygen atom (**Fig. 6**).

NADPH dependence for *C. glabrata* CBS138 CR was interpreted by aligning the crystallographic structure of *S. salmonicolor* AKR protein together with that of *C. glabrata* and followed by comparing the cofactor attachment domain between the two proteins(**Fig. 7**). The cofactor domain mapping of the target protein together with comparing it with that of *S. salmonicolor* (1UJM) and *S. cerevisiae* (YDR541c, YGL157w and YOL151w) AKR proteins has been accomplished using standard alignment and labeling tools of PyMOL (www.pymol.org).

Thus homology model construction, alignment of important amino acids residues, analyzing substrate binding site, exploring cofactor domain and docking score comparisons with other standard ALR/ CR group of enzymes suggested that XP_445913.1 should belong to the same family of enzymes as the standards and thus we treated XP_445913.1 as CR (Carbonyl Reductase) group of protein from *Candida glabrata* CBS138. These findings encouraged us to try and explore its actual biocatalytic potential in realistic experimental condition. Thus we cloned the gene in heterologous vector, over-expressed and subsequently purified the protein, characterized through kinetic studies, calibrated through pH and temperature and finally exploited it in actual bioconversion through trinomial bioengineering as described in our earlier report.

Optimization of Bioconversion

The optimization of the reaction system was accomplished by analyzing the outputs with any two variables of the three key factors controlling the bioconversion such as monophasic system, biphasic system and NADPH as cofactor. Apparently, best yields were obtained when biphasic reaction system was used together with cofactor regeneration as reported by other researchers.¹⁻⁵ As mentioned earlier¹⁰, gene encoding Glucose dehydrogenase (GDH) from *Bacillus subtilis* has been cloned in the recombinant microbial cell factory together with CR from *Candida glabrata* CBS 138 in order to follow cofactor regeneration of NADPH from oxidized NADP⁺ via exogenously added substrate glucose. When whole cells over-expressing the CR and GDH proteins (cofactor regenerating) were employed with previously reported conditions¹⁰ within biphasic reaction system, the substrate being in butyl acetate phase while recombinant host in the buffer phase, the product formation improved rapidly up to 4h and obtained steady state within 6h (88.3% bioconversion from COBE to CHBE, **Fig. 8A**). In contrast, without cofactor regeneration even in biphasic system, only 5% bioconversion was achieved together with having a longer reaction time ~ 8--10 h (**Fig. 8A**). However, a cofactor regenerating system within single buffer phase (0.1 M Potassium phosphate buffer, pH 7.5) produced a bioconversion of 20.80% necessitating the contribution of cofactor in enhancing the product yield. **Fig.8B** depicts a comparative portfolio of bioconversion under three aforementioned conditions.

DNA shuffling- Library formation and mutant selection for future host construction

We employed error-prone PCR (ep-PCR) in our previous study¹⁰ to construct highly stress tolerant mutant as our biocatalysis host and thus achieved significant improvement of product yield with elongation of cell fitness during the biphasic reaction. However we are still in search for further improvement of cell phenotype as we believe that cell fitness can be a key factor to tower the product yield during such bioreaction. Many such strategies are under investigation, among them DNA shuffling has emerged as a successful tool to improve cell phenotype in multitude of conditions.¹⁷⁻²⁰

DNA shuffling was carried out with a little modification of the procedure as described by Stemmer, 1994²¹ taking whole *crp* operons of three ep-PCR mutants (M1 ~ M3) as templates. The acquisition of M1~

M3 and their amino acid mutations have been already described in our previous publication.¹⁰ 4 µg of total template DNA was used for DNA shuffling. The total template DNA was digested by DNase-I at 15°C for 3 mins and subsequently 50--200 bp DNA fragments had been recovered from gel electrophoresis for further process. Afterwards, DNA fragments were subjected to undergo PCR without primer and finally chimeric *crp* was recovered by amplification with forward and reverse *crp* primer, the sequences as discussed in our previous publication.

The DNA shuffling library was finally constructed by cloning the chimeric *crp* into pACYC Duet-I plasmid with *Kpn* I and *Bam* HI and subsequent introduction into *E. coli* DH5α as discussed earlier¹⁰. The introduction of chimeric *crp*-pACYC Duet-I conjugate into *E. coli* DH5α through electroporation yielded a variant library in the order of 10⁴~10⁵. DNA shuffling mutants were grown in LBGMg medium (Bacto tryptone 10 g/L, Yeast extract 1g/L, NaCl 10 g/L, Glucose 2g/L, MgSO₄ 10 mM) and the winner was selected by subjecting the mutant library in three rounds of selection with 0.23~0.25% (v/v) toluene. 50 mL High Density Polypropylene centrifuge tubes (BD Biosciences, USA) with parafilm sealing have been used for culturing the variant with proper oxygenation (37°C, 200 rpm). Proper dilution of the culture from each round was plated onto LBGMg-agar plates to isolate the colonies. Selected clones from the third round were taken, the plasmid DNA was isolated by mini-preparation with QIAGEN plasmid isolation kit (Qiagen, USA). The shuffled mutations were verified by DNA sequencing and relevant plasmids were re-introduced into fresh *E. coli* DH5α background to create fresh variants. The fresh variants were challenged in LBGMg medium under 0.4% - 0.5% (v/v) toluene pressure to select the variant with best growth profile.

One DNA shuffling mutant (DSM) was isolated and sequenced to reveal amino acid substitutions such as T127N F136I T208N. The DSM revealed better growth profile against higher concentrations of toluene (0.4% - 0.5% (v/v)). Under 0.40% toluene, the DSM reached to OD₆₀₀ ~ 2.5 in 24 h where other mutants' growth (M1~ M3) remained within OD 2.0 ~ 2.3 (**Fig. 9A**). In 0.50% toluene, where other reached a saturation OD ~ 2.0 in 22 h, DSM reached a OD ~ 2.4 in the same time (**Fig. 9B**).

Discussion and Conclusion

The CR from *Candida glabrata* was identified through BLASTP guided sequence search method. It is known that if there is a significant sequence match between two proteins, then it is highly likely that there are similarities in their functions. In this research work, it was found that there was a 55% ~ 62% identity match between CR and three reported AKR family of proteins YDR151c, YGL157w and YOL151w.¹¹ This suggests that the target protein might have an AKR like activity. The overlapping of active site between the target and the template proteins also imply similar tertiary structure of the proteins.

Since the CR has been structurally elusive together with YDR151c, YGL157w and YOL151w, all the proteins have been modeled to churn any similarities between their structures as well as the active sites. The models were first validated then the active site comparison revealed that similar hydrophobic channels spanned the target and template protein residues including structurally elucidated proteins 1UJM and 1Y1P. Thus it is a plausible assumption the substrate COBE may bind with the proteins in the same way, however the specificity and affinity may change depending upon the hydrophobicity of the channel. Furthermore, an *in silico* docking of COBE with the proteins revealed that COBE has even more binding affinity towards CR than other template proteins. This really encouraged us to try the CR from *Candida glabrata* in realistic bioconversion.

One critical bottleneck for enzymatic bioconversion is the hydrophilicity of the enzyme since enzymes usually are dissolved in buffered system for prolonging its stability and activity. Thus, a prior investigation of CR in this sort through GRAVY (grand average of hydropathicity) analysis (<http://web.expasy.org/cgi-bin/protparam/protparam>) revealed a negative gravity index as -0.359. This indicated protein's hydrophilicity²² thus advocating its competence in buffer mediated bioconversion. Then we challenged the enzyme in actual bioconversion and optimized it through whole cell biotransformation.

It is important to note that biphasic biotransformation provides better output in this reaction. The reason being that improved solubility of COBE in the organic phase (compared to aqueous) enhances the

substrate carriage to the host cells. Also, COBE undergoes partial hydrolysis in aqueous phase which retards the rate of reaction in the monophasic biotransformation (data not shown). Cofactors are required to supply the hydrogen for the biotransformation.^{1-5,23}

Harboring the aforementioned highly efficient recombinant enzyme for improved biotransformation, we have tried to overcome secondary challenges in biphasic biocatalysis such as tolerance of the host cell towards organic phase and improved substrate uptake inside the host cell reported in our previous publication.¹⁰ We have demonstrated that error prone PCR has been undertaken to improve cell phenotype during biocatalysis. Also, cell phenotype has been improved by DNA shuffling. For DNA shuffling we have used whole *crp* operon from the mutants because the *crp* operon contains three parts: the transcription factor binding site (Transcription Factor B involving two CRP-cAMP binding sites, four FISbinding sites, and three *crp* native promoters), the *crp* gene, and a specifically designed *rrnB* terminator.^{24,25} The resultant *in vitro* recombination yielded a super-performing mutant (DSM) which showed improved performance than the mothers against extremely hydrophobic toluene. Toluene has been selected as the challenging solvent because a “winner” against such extremely hydrophobic solvent would have good chance to survive other organic solvents too. LBGMg medium has been a choice of medium for survival of colonies against organic solvent.²⁶ The DSM combines mutations at three regions namely T127 (in the cAMP binding pocket- α helix that stabilizes cAMP- CRP binding)²⁷, F136 (stabilizing interdomain hinge)²⁸ and T208 (in the DNA binding domain).²⁷ We propose that implementation of DNA Shuffling mutant in future biotransformation might improve the product titer even better and might gain other applications as well.

Acknowledgement

The authors are highly thankful to School of Chemical and Biomedical Engineering (SCBE), Nanyang Technological University (NTU), Singapore for providing the laboratory infrastructure to carry out biomolecular experiments. The authors are also grateful to Dr. Jiang Rongrong, Asst/P. SCBE, NTU, Singapore for her help and suggestions regarding the study.

REFERENCES

1. Wang Q, Shen L, Ye T, Cao D, Chen R, Pei X et al. Overexpression and characterization of a novel (S)-specific extended short-chain dehydrogenase/reductase from *Candida parapsilosis*. *Bioresour Technol* 2012; 123:690-694
2. Cao H, Mi L, Ye Q, Zang, G, Yan M, Wang Y et al. Purification and characterization of a novel NADH-dependent carbonyl reductase from *Pichia stipitis* involved in biosynthesis of optically pure Ethyl (S)-4-Chloro-3-hydroxybutanoate. *Bioresour Technol* 2011; 102: 1733--1739.
3. Ye Q, Cao H, Mi L, Yan M, Wang Y, He Q et al. Biosynthesis of (S)-4-chloro-3-hydroxybutanoate ethyl using *Escherichia coli* co-expressing a novel NADH-dependent carbonyl reductase and a glucose dehydrogenase. *Bioresour Technol* 2010; 101: 8911--8914.
4. You ZY, Liu ZQ, Zheng YG. Characterization of a newly synthesized carbonyl reductase and construction of a biocatalytic process for the synthesis of Ethyl (S)-4-Chloro-3-hydroxybutanoate with high space-time yield. *Appl Microbiol Biotechnol* 2014; 98: 1671--1680.
5. Liu Y, Xu Z, Jing K, Jiang X, Lin J, Wang F et al. Asymmetric reduction of ethyl 4-chloro-3-oxobutanoate to ethyl (R)-4-chloro-3-hydroxybutanoate with two co-existing, recombinant *Escherichia coli* strains. *Biotechnol Lett* 2005; 27: 119--125.
6. Chen XH, Wei P, Wang XT, Zong MH, Lou WY, A novel carbonyl reductase with anti-Prelog stereospecificity from *Acetobacter sp.* CCTCC M209061: purification and characterization. *PLoS One* 2014; 9: e94543.
7. Wang L-J, Li C-X, Ni Y, Zhang J, Liu X, Xu J-H. Highly efficient synthesis of chiral alcohols with a novel NADH-dependent reductase from *Streptomyces coelicolor*. *Bioresour Technol* 2011; 102: 7023--7028.
8. Yoon SA, Kim HK. Development of a bioconversion system using *Saccharomyces cerevisiae* Reductase YOR120W and *Bacillus subtilis* glucose dehydrogenase for chiral alcohol synthesis. *J Microbiol Biotechnol* 2013; 23: 1395--1402.
9. Lee DH, Lee YJ, Ryu YW, Seo J H, Molecular cloning and biochemical characterization of a novel erythrose reductase from *Candida magnoliae* JH110. *Microb Cell Fact* 2010; 9.
10. Basak S, Ghosh SK, Punetha VD, Aphale AN, Patra PK, Sahoo NG. An experimental modeling of trinomial bioengineering- crp, rDNA, and transporter engineering within single cell factory for maximizing two-phase bioreduction. *Int J Biol Macromol* 2017; 95: 818--825
11. Kaluzna IA, Matsuda T, Sewell AK, Stewart JD. Systematic investigation of *Saccharomyces cerevisiae* enzymes catalyzing carbonyl reductions. *J Am Chem Soc* 2004; 126:12827-12832
12. Melo F, Feytmans E 1998. Assessing protein structures with a non-local atomic interaction energy. *J Mol Biol* 1998; 277: 1141--1152.

13. Benkert P, Tosatto SCE, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics* 2008; 71:261-277
14. Christen M, Hünenberger PH, Bakowies D, Baron R, Bürgi R, Geerke DP, Heinz TN, Kastenholz MA, Kräutler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF. The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem* 2005; 26: 1719--1751
15. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993; 26: 283--291
16. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 2006; 22: 195--201
17. Zhang H, Chong H, Ching CB, Jiang R. Random mutagenesis of global transcription factor cAMP receptor protein for improved osmotolerance. *Biotechnol Bioeng* 2012; 109: 1165--1172
18. Zhang H, Chong H, Ching CB, Song H, Jiang R. Engineering global transcription factor cyclic AMP receptor protein of *Escherichia coli* for improved 1-butanol tolerance. *Appl Microbiol Biotechnol* 2012; 94: 1107--1117
19. Cramer A, Raillard SA, Bermudez E, Stemmer WP. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* 1998; 391: 288--291
20. Gao X, Zhao H, Zhang G, He K, Jin Y. Genome shuffling of *Clostridium acetobutylicum* CICC 8012 for improved production of acetone-butanol-ethanol (ABE). *Curr Microbiol* 2012; 65: 128--132
21. Stemmer, WPC. DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA*, 91: 10747--10751
22. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982; 157:105-132
23. He Y-C, Tao Z-C, Zhang X, Yang Z-X, Xu J-H. Highly efficient synthesis of ethyl (S)-4-chloro-3-hydroxybutanoate and its derivatives by a robust NADH-dependent reductase from *E. coli* CCZU-K14. *Bioresource Technol* 2014; 161: 461--464
24. Brosius J, Dull TJ, Sleeter DD, Noller HF. 1981. Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *Journal of Molecular Biology* 1981; 148:107-127.
25. Okamoto K, Freundlich M. 1986. Mechanism for the autogenous control of the *crp* operon: Transcriptional inhibition by a divergent RNA transcript. *Proceedings of the National Academy of Sciences USA* 1986; 83:5000-5004.
26. Basak S, Song H, Jiang R. Error-prone PCR of global transcription factor cyclic AMP receptor protein forenhanced organic solvent (toluene) tolerance. *Process Biochem* 2012; 47: 2152--2158
27. Won HS, Lee YS, Lee SH, Lee BJ. Structural overview on the allosteric activation of cyclic AMP receptor protein. *Biochim Biophys Acta* 2009;1794:1299-1308.

28. Fas BA, Tutar Y, Haliloglu T. Dynamic Fluctuations Provide the Basis of a Conformational Switch Mechanism in Apo Cyclic AMP Receptor Protein, PLoS Comput. Biol. 2013; 9: e1003141.

Accepted Manuscript

Table 1. Comparison of Binding site

	Catalytic residues			Hydrophobic channel					
<i>C. glabrata</i>	S133	Y 175	K 179	F 94	I 172	F 245	V229	P 241	-
<i>S. salmonicolor</i>	S134	Y 177	K 181	F 97	L 174	L 241	W 226	A 238	P170
<i>S. cerevisiae</i> YDR541C	S129	Y 167	K 171	Y 89	I 164	P 235	-	-	-
<i>S. cerevisiae</i> YGL157W	S131	Y 169	K 173	F 91	V 166	Y 239	I223	-	-
<i>S. cerevisiae</i> YOL151W	S127	Y 165	K 169	-	V 162	I 232	-	-	P161

Table 2. Docking scores of proteins

Protein	Binding affinity (Kcal/mol)
<i>C. glabrata</i> XP_445913.1	-4.6
<i>S. cerevisiae</i> YDR541c	-3.9
<i>S. cerevisiae</i> YGL157w	-3.5
<i>S. cerevisiae</i> YGL157w	-4.5
AKR protein from <i>S. salmonicolor</i>	-3.7

Accepted Manuscript

C. glabrata 1 MASDNS----NTTVFVSGATGFIAQHVVVRQLLDQNYKVI GSVRSAEKGDHINKNVIFK--G
S. salmonicolor 1 AKIDNAVLPEGSIVLVTPGANGFVASHVVEQLLEHGKVRGTARSASKLANLQKRWDAKYP
YDR541c 1 MS-----NTVLVSGASGFIALHILSOLLKQDYKVI GTVRSHEKEAKLRFQH--N
YGL157w 1 MTT-----DTTVFVSGATGFIALHIMNDLLKAGYTVIGSGRSQEKNDGLLKRENN--N
YOL151w 1 M-----SVFVSGANGFIAQHIVDILLKEDYKVI GRSRQEKAEINITEAFGN--N

C. glabrata 55 GDFNYEIVKDISDETFADHVEEKHGKDIKIVLHTASPFHENTTDIEKDLLIPAVNGTKGI
S. salmonicolor 61 GRRETAVVEEXLKGAYDEVIKGA---AGVAHIASVVSFSN--KYDEVVTPAIGGTLNA
YDR541c 50 FNLITLIVPDISHFNAFDKVLQKRGREIRYVLTASPPHYDTTEYEKDLLIPALEGTRNI
YGL157w 52 PKLSMEIVEDIAAFNAFDEVKKGKGIKIVLHTASPFHETTFNEKDLLIPAVNGTKSI
YOL151w 48 PKFSMEVVPDISKLLDAFDHVEQKHGKDIKIVLHTASPFCEDTTDSERDLLIPAVNGVGI

C. glabrata 115 LESIKKYAAQTVERVVVTSFAANTSIVDMFYAKLSKKTITEESWNQDTWESCC-----
S. salmonicolor 115 LRAA--AATPSVKRFVLTSSIVSALI-FK---PVEGTYLDEKSNWLESIDKAKTLPESD
YDR541c 110 LNSIKKYAADTVERVVVTSCTAITLAK---MDFSVVFTTEESWNEATWESCC-----
YGL157w 112 LEAIKKYAADTVEKVI VTSSTAALVPTD---MNGKDTVITEESWNKDTWDSCC-----
YOL151w 108 LHSIKKYAADSVERVVLTSYAAVDFMAK---ENKKSITFNESWNPATWESCC-----

C. glabrata 169 -SDPTIRGYCGSKKFAEKAAWDFYNANKDSVKFKLSIINEVYVFGPQNYVFPGKKILNTSS
S. salmonicolor 169 PQKSLWVFAASKTEAFLAAWKFXDENKPE--HFTTNAVLENYITIGT-IFDFETQSG--STS
YDR541c 161 -IDGYNAYDASKKFAEKAAWDFYKFNEDHFKKLTIVNPSLFGPOLFDDEDVHGHNTSC
YGL157w 163 -ANAVAYCGSKKFAEKTAWEFLKENKSVKFTLSTINFGVFGPQMFADSLKHGINTSS
YOL151w 159 -SLFVNAYCGSKKFAEKAAWDFLENRDSVKFELTAVNVEVYVFGPQMFQKDVKKHLNTSC

C. glabrata 228 EVINSLVHLKKDDPLPEF-AGGHIDVRDVAKAHILAFQKDELIEQRLMLHAGLETTQTLL
S. salmonicolor 224 GWXXSIFNGEVSPALALXPPQYVSAVDIGLLHLGCLVLPQIERRRVYGTAGCFDWNITVL
YDR541c 220 EMINGLIHTFVNASVPDF-HSIFIDVRDVLAHLVAFOKENTACKRLVVTNGKFGNODIL
YGL157w 222 GIVSELIHSKVGGEFYNY-CCPFI DVRDVSKAHLVAIEKPECTGQRLVLSGELFCQOEIV
YOL151w 218 ELVNSLMHLSPEDKIPEL-FGGYIDVRDVAKAHLVAFQKRETIQQLVISEARFTMQDVL

C. glabrata 287 DIINECFEELKCKLPAGKPGTGNPDD-AL-TPVDNSKTKKLL---G-FEFDLKKDLYDT
S. salmonicolor 284 ATFRKLYPSK--TFPADEFDQGDLDL---SKFDTAPSLEILKSLGRPCWRSITEESIKDL
YDR541c 279 DILNEDFQLRGLIPLGKPGTGQVIDRG-STTDNSATRKIL---G-FEFRSLHESVHDT
YGL157w 281 DILNEEFELKCKIATGEPATGSPFLEKNSCKFNSKTKKLL---G-FEYNLKDCIVDT
YOL151w 277 DILNEDFVVLKENIPVCKPGSGATHNTLG-ATLDNKKSKLL---G-FRFRNLKETIDDT

C. glabrata 341 ISQILEAEKRSN
S. salmonicolor 337 VG----S-ETA
YDR541c 334 AAQILKKE-NRL
YGL157w 337 AAQMLEVQ-NEA
YOL151w 332 ASQILKKE-GRI

Figure 1. Multiple sequence alignment of carbonyl reductase from *Candida glabrata* CBS138 with other reported Aldo-keto Reductase group of proteins namely one aldose reductase from *Sporobolomyces salmonicolor* and three *Saccharomyces cerevisiae* derived genes encoding carbonyl reductases (*YDR541c*, *YGL157w*, *YOL51w*). The conserved domains are highlighted as black and grey boxes

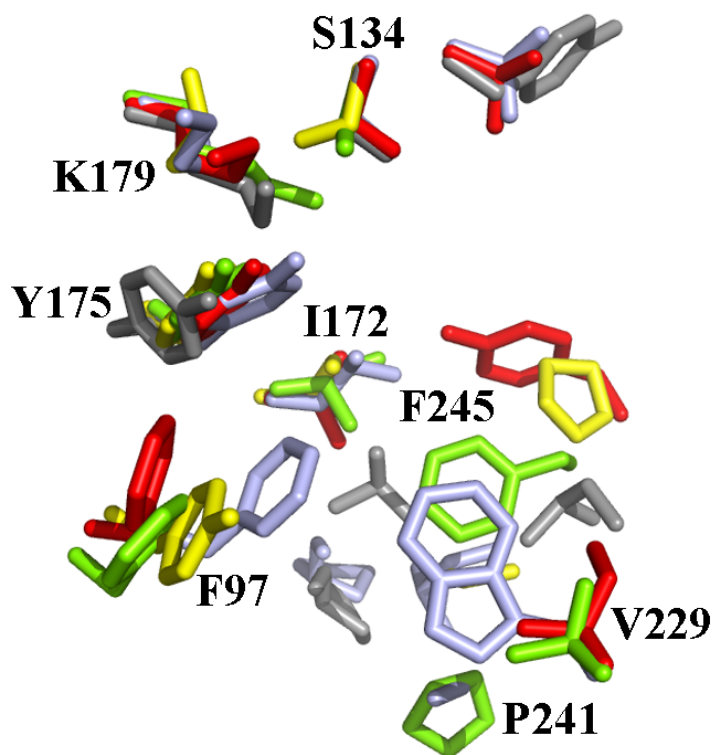


Figure 2. Positions of important amino acids for COBE binding shown in stick representation by pyMOL. Overlap of crystal structure from *S. Salmonicolor* (light blue) and homology model structure from *C. glabrata* (green), *S. cerevisiae* YDR541C(yellow), *S. cerevisiae* YGL157W(red), *S. cerevisiae* YOL151W(gray).

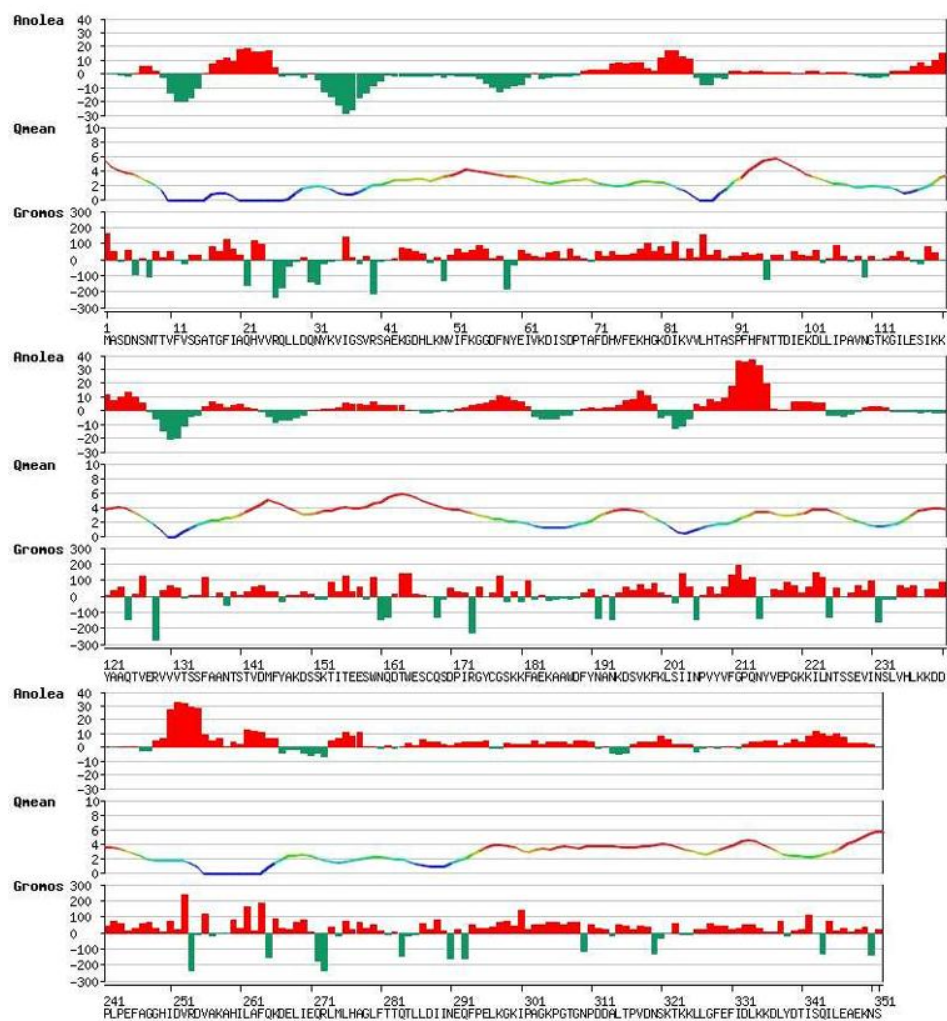
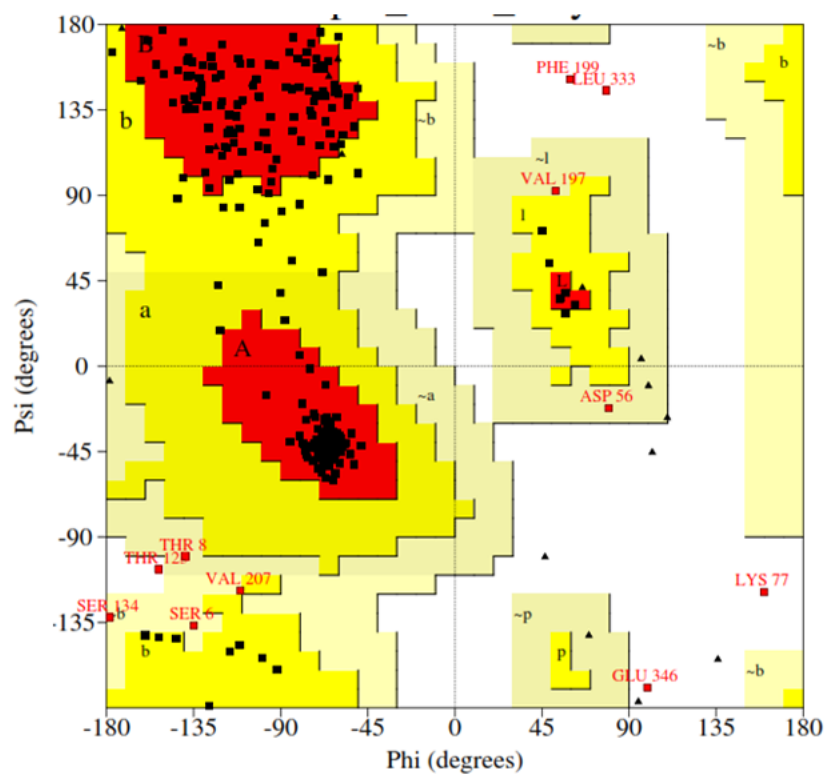


Figure 3. ANOLEA, QMEAN and GROMOS plot of CR from *Candida glabrata* CBS138.



Plot statistics

Residues in most favoured regions [A,B,L]	271	86.0%
Residues in additional allowed regions [a,b,l,p]	33	10.5%
Residues in generously allowed regions [-a,-b,-l,-p]	6	1.9%
Residues in disallowed regions	5	1.6%
Number of non-glycine and non-proline residues	315	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	21	
Number of proline residues	14	
Total number of residues	352	

Figure 4. Ramachandran Plot of CR from *Candida glabrata* CBS138.

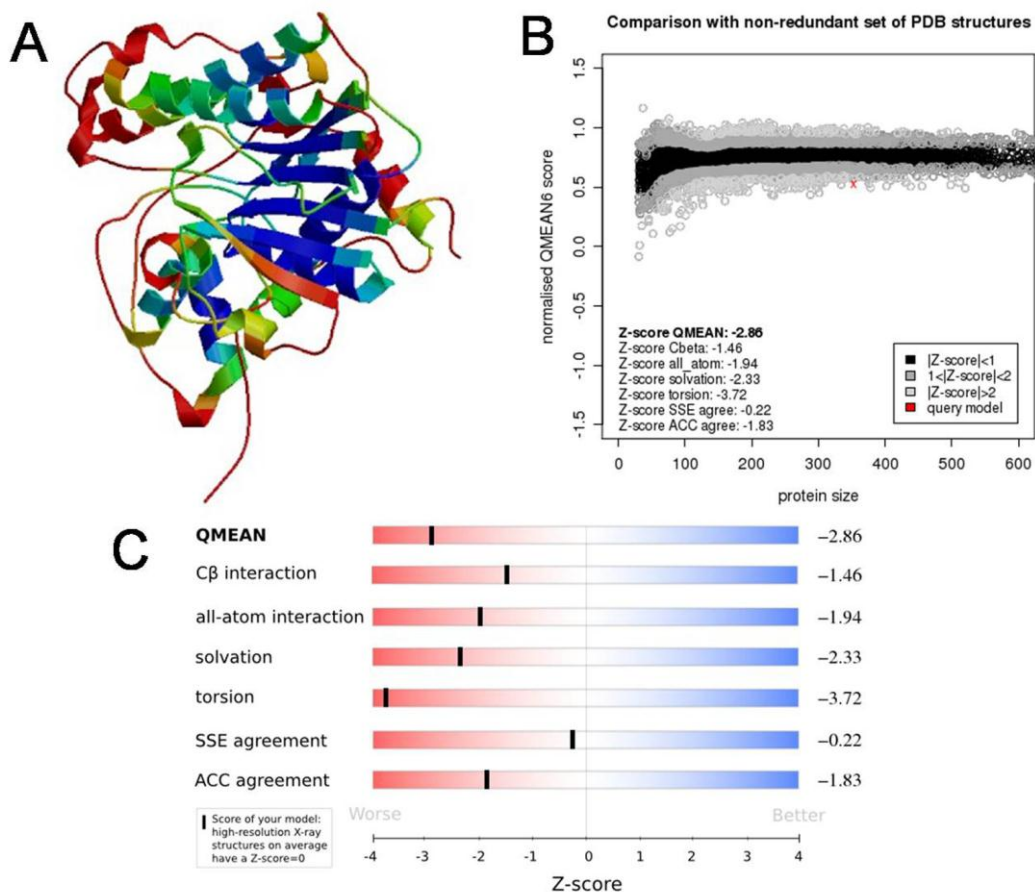


Figure 5. Different model validation parameters of CR from *Candida glabrata* CBS138. **A.** Coloring residue plot with errors (blue depicting most reliable residues while red suggests potentially unreliable residues). **B.** Pseudo-energy plot of contributing terms with their z -scores. **C.** QMEAN6 plot with query residues.

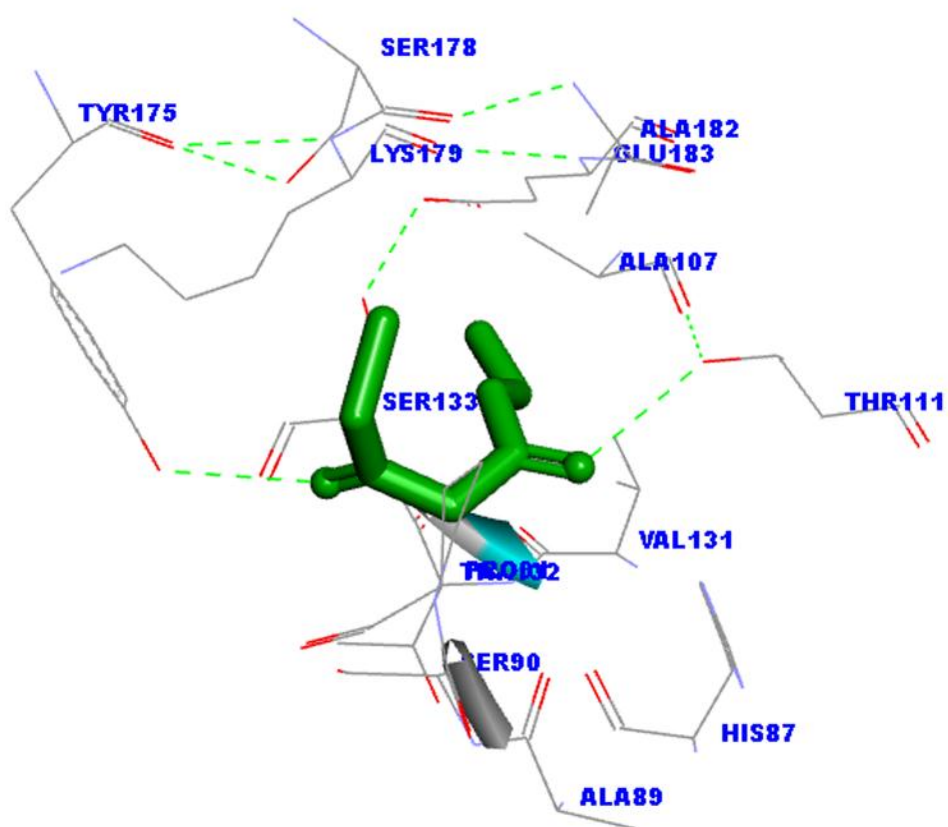


Figure 6. Binding site analysis of COBE in CR from *Candida glabrata* CBS138. The binding site analysis has been done using the standard protocol in Discovery Studio Visualizer 3.5. The dotted bond represents hydrogen bonding.

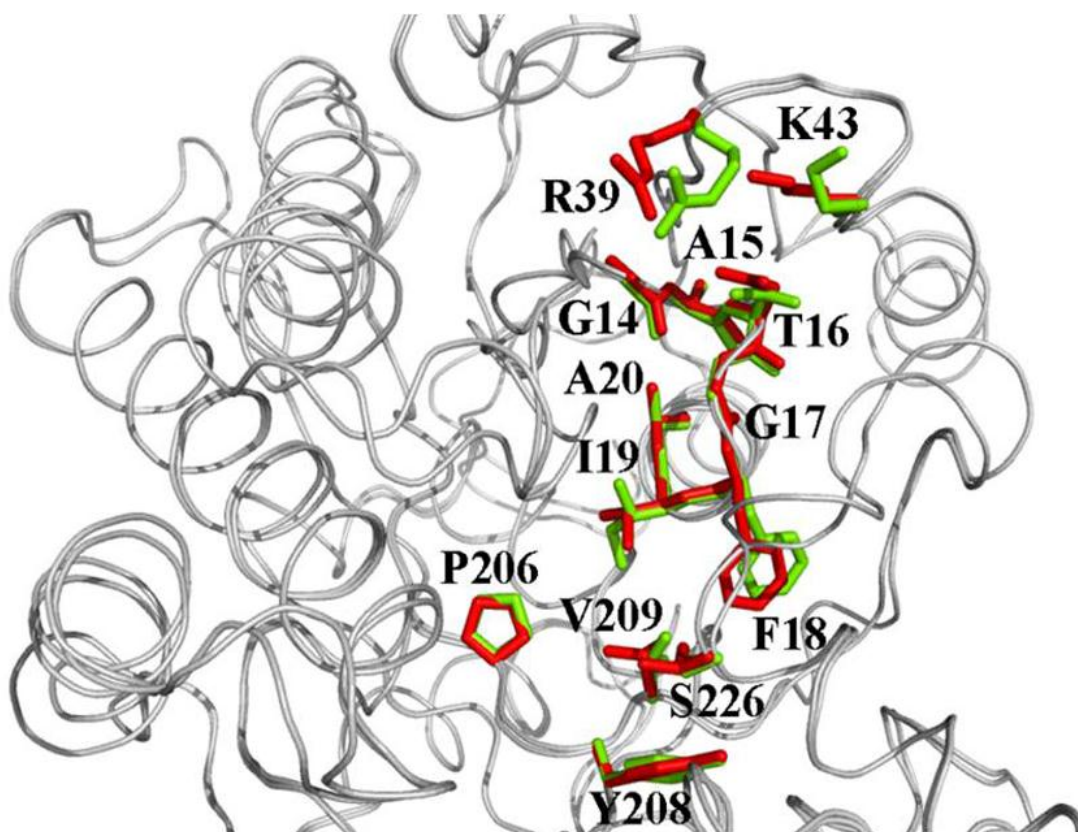


Figure 7. Positions of important amino acids for NADPH binding for *C. glabrata* (green) and *S. Salmonicolor* (red).

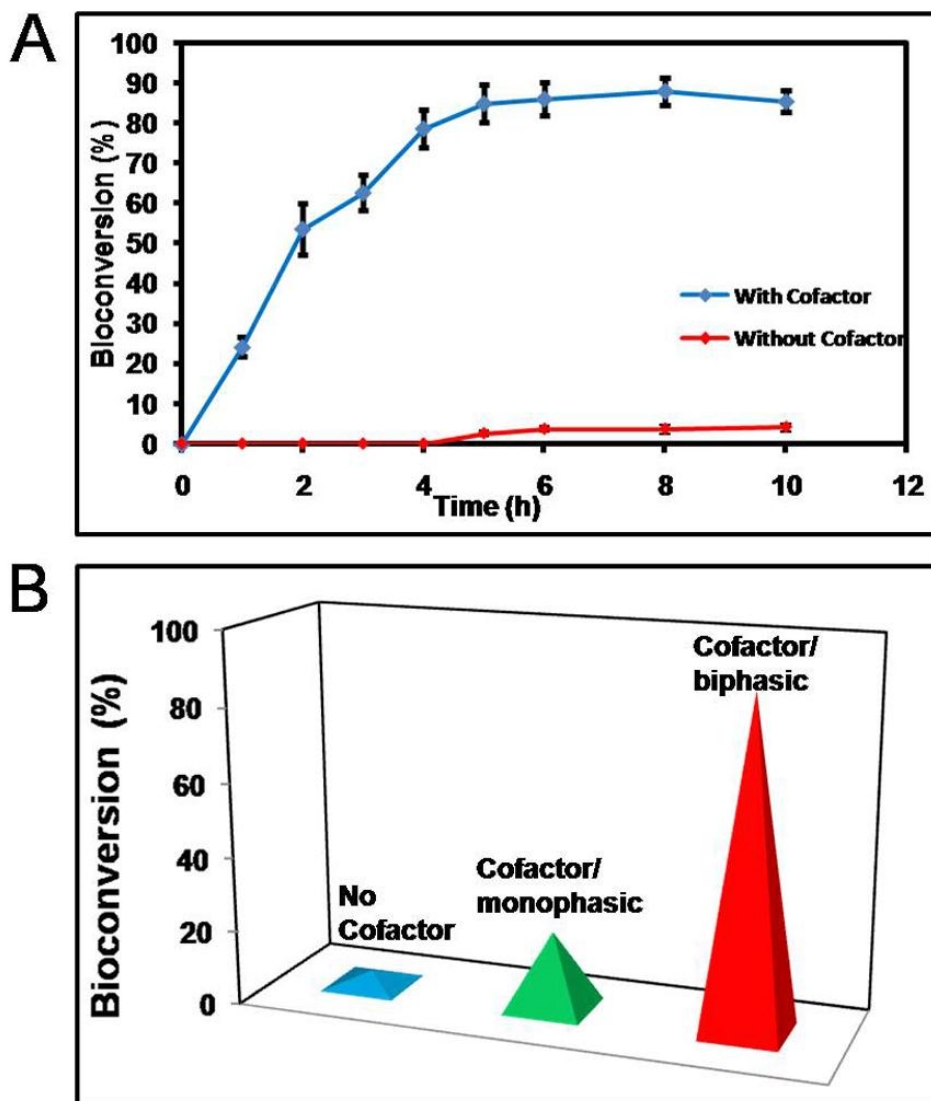


Figure 8. (*R*)-CHBE production profile in biotransformation reaction. **A.** Yield profile with or without cofactor regeneration **B.** Comparative study of bioconversion under three reaction schemes The data points are the average (\pm standard deviation) of three independent observations ($n = 3$).

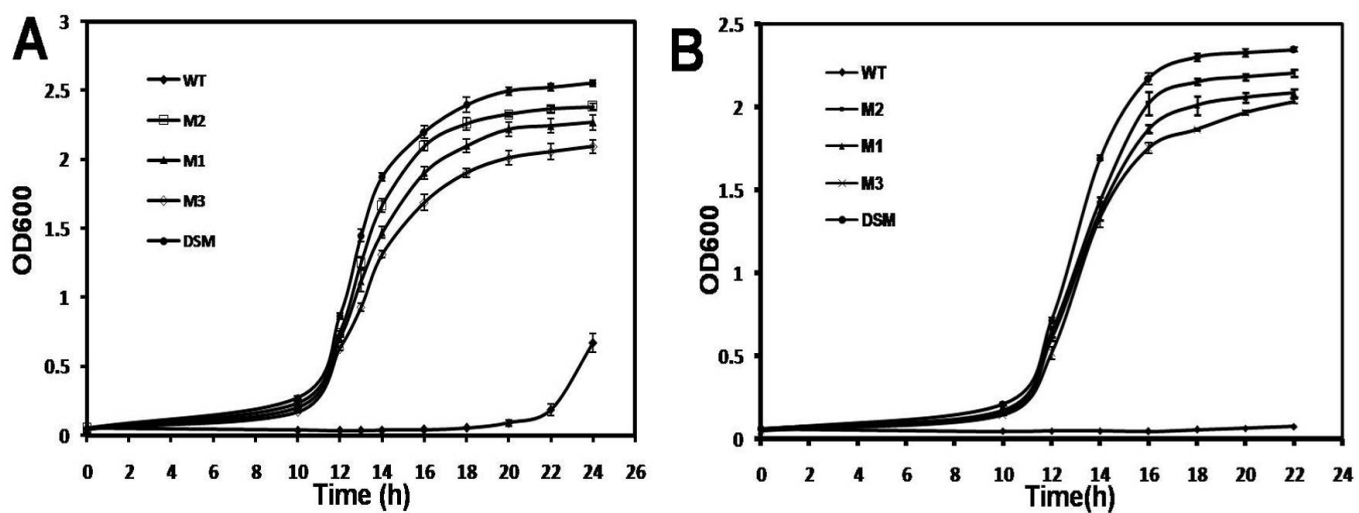


Figure 9. Tolerance profile of DNA shuffling mutant (DSM) with respect to parent mutants. The tolerance has been analyzed by growth of the mutant against hydrophobic solvent such as **A.** 0.40% (v/v) toluene **B.** 0.50% (v/v) toluene