Technical Progress Report

for

The Biocatalytic Desulfurization Project

Quarterly Report

Report Start Date: 03/20/2003
Report End Date: 06/19/2003

Principal Authors:
Steven E. Bonde, Petro Star Inc.
David Nunn, Diversa Corporation

Report Issue Date: July, 2003

DOE Award Number: DE-FC26-02NT15340

Submitting Organizations:

Petro Star Inc.
3900 C Street
Suite 401
Anchorage, AK 99503

Diversa Corporation
4955 Directors Place
San Diego, CA

Disclaimer

Table of Contents

Abstract

The analysis of Petro Star diesel sulfur species is complete and a report is attached. Further analytical efforts will concentrate on characterization of diesel fuel, hydrodesulfurized to varying degrees, in order to determine sulfur species that may be problematic to hydrogen treatment and represent potential target substrates for biodesulfurization in a combined HDS-BDS process. Quotes have been received and are being considered for the partial treatment of Petro Star Inc. marine diesel fuel.

Direction of research is changing slightly; economic analysis of the hyphenated – BDS-HDS, BDS-CED – has shown the highest probability of success to be with a BDS-HDS process where the biodesulfurization precedes hydrodesulfurization. Thus, the microorganisms will be tailored to focus on those compounds that tend to be recalcitrant to hydrodesulfurization and decrease the severity of the hydrodesulfurization step. A separate, detailed justification for this change is being prepared.

Research activities have continued in the characterization of the desulfurization enzymes from multiple sources. Genes for all DszA, -B, -C and -D enzymes (and homologs) have been cloned and expressed. Activity determinations, on a variety of substituted benzothiophene and dibenzothiophene substrates, have been carried out and continue. In addition, chemical synthesis efforts have been carried out to generate additional substrates for analytical standards and activity determinations.

The generation of a GSSM mutant library of the *Rhodococcus* IGTS8 *dszA* gene has been completed and development of protocols for a high throughput screen to expand substrate specificity are nearing completion. In an effort to obtain improved hosts as biocatalyst, one hundred-thirty *Rhodococcus* and related strains are being evaluated for growth characteristics and other criteria deemed important for an optimal biocatalyst strain. We have also begun an effort to generate derivatives of the entire IGTS8 BDS plasmid that will allow for its easy transfer and manipulation into a variety of hosts. To support this activity and to gain an understanding of additional genes that may potentially affect BDS activity, the nucleotide sequence of the entire complement of plasmids in IGTS8 is being determined.

Lastly, we continue to develop genetic screens and selections for the discovery and improvement of the biodesulfurization genes and strains.

1.0 Executive Summary

Task 1, 8

Analysis of Petro Star diesel has been completed. The methodology developed will be used with any future fuels.

Tasks 2, 9

Cloning and expression of all available *dsz* genes has been completed. Enzymatic activities on a variety of benzothiophene and dibenzothiophene substrates have been qualitatively determined. A number of enzymes have been purified and specific activities have been determined on a number of substrates. Analytical assays for all substrates using mass spectrometry have been developed to support enzyme characterization and directed evolution efforts. A GSSM library of DszA variants (where every position in the protein is mutated to every other amino acid) has been generated and undergone quality control to insure completeness. Multivariant analysis and refinement of an *in vitro* high throughput screen for variants with improved activity on substituted $DBTO_2$ and BTO_2, relative to activity on $DBTO_2$, is being carried out.

Task 3

In order to develop a biocatalyst with improved activities on high molecular weight DBTs and BTs an effort to characterize >130 Rhodococcus and related strains was undertaken. Cultures were obtained from both the ATCC and DSMZ and were characterized for growth rates on various rich and minimal medias, aeration requirements, and ease of manipulation on solid media. These criteria were used to identify a subset of candidate strains for introduction of the *Rhodococcus* IGTS8 BDS plasmid and characterization of BDS activity in comparison to IGTS8. In support of this activity the entire nucleotide sequence of the BDS plasmid is being determined in order to identify additional genes that may be important for BDS as well as to aid in the generation of derivatives for easy transfer to a variety of bacterial strains.

Task 4

Efforts in this area, to expand substrate specificities to non-thiophenic sulfur species, have been minimized as these compounds do not appear to represent significant challenges for hydrodesulfurization. They would not be important targets for a biocatalyst in a combination process with HDS.

2.0 Experimental

Bioanalytical methods: A revised method has been developed for analysis of BT and DBT compounds All samples were injected from a 96-well plate format using an *HTS*Pal auto- sampling robot (Leap Technologies, Carrboro, NC) into various mixtures of H2O/acetonitrile (0.1% formic acid), delivered by Shimadzu (Kyoto, Japan) LC-10ADvp pumps at various flow rates, and passed through a Zorbax Eclipse XDB-C8, 4.6

x 150 mm, 5um liquid- chromatography column (Agilent, Palo Alto, CA) and into an API 4000 triple-quad mass spectrometer (Applied Biosystems, Foster City, CA) where electrospray ionization (ESI) and multiple reaction monitoring (MRM) was performed in the positive ion mode. Instrumentation control and data generation was accomplished using Analyst 1.2 software (Applied Biosystems, Foster City, CA).

Molecular biology. Genomic DNAs were prepared using standard methods (1). PCR amplification of sequences to be cloned was carried out using either genomic DNA or purified plasmid as template and *Pfu* polymerase (Stratagene, La Jolla, CA). Optimized amplification conditions were determined by varying incubation temperatures and numbers of cycles. PCR products of the coding regions of DszA, DszB and DszC homologues were amplified by PCR and cloned into pASK5 plasmid between *Nco*I and *Bgl*II sites to create a plasmid that expresses native or carboxy-terminal hexahistidine-tagged oxidoreductase. For amino-terminal hexahistidine-tagged proteins, the PCR product was cloned into pASK1 between *Bgl*II and *Hind*III sites.

For construction of the *Rhodococcus dszA* GSSM library, the *dszA* gene,cloned into pASK5,was mutagenized at every position in the protein using degenerate oligonucleotides where the residue codon to be modified was represented by the sequence NNK (N= G, A, T or C; K=G or C) and was carried out by Diversa proprietary GSSM technology. Quality control of the GSSM library was carried out by the random selection of codon mutants and nucleotide sequencing to determine the extent of the substitutions obtained.

For sequencing of the *Rhodococcus* IGTS8 plasmids, a procedure for the isolation of plasmids from this organism was developed. A single colony was inoculated into 200mls of 2xYT medium in a 2L flask and incubated for 48 hours at 30^{o}C, 250 RPM. Fifty mls of culture was then inoculated into each of four flasks containing 400ml 2xYT with 2% glycine and further incubated for 8 hours. Cells were harvested, washed with 0.5X volume of 50mM EDTA, pH8.0, and the cell pellet resuspended in 80 ml Qiagen solution I. Mutanolysin (100U/ml) and lysozyme (2mg/ml) was then added and the cells incubated for 1 hour at 37^{o}C. From this point, directions for the Qiagen Large Construct Kit were followed.

To construct recombinant clones for sequencing,IGTS8 plasmid DNA was partially digested with Sau3AI and 1-2kb fragments were agarose gel-purified. These fragments were then ligated into *Bam*HI-cut pZero2 (Invitrogen, Carlsbad, CA) and recombinants selected by plating on L-agar with ampicillin. Plasmid DNA for sequencing was prepared from each recombinant using a HTP microtiter plate format.

Expression and Purification. *E. coli* TOP10 cells (Invitrogen, Carlsbad CA) were used as a host to express recombinant oxidoreductases. Freshly transformed colonies were inoculated into LB medium with kanamycin. 1 L cultures were grown until OD_{600}~1 and

expression was induced with anhydrotetracycline (10-100 µg/l) and carried on overnight at 30°C. Cells were lysed in buffer I (50 mM sodium phosphate; 100mM NaCl supplemented with lysozyme (Epicentre, Madison, WI) by French-pressure treatment. Cell lysates were clarified by centrifugation at 15,000xg for 30 min and recombinant protein was bound to pre-equilibrated Ni^{+2}-NTA resins at 4°C for 1 hr. Resins were washed with buffer I containing imidazole (20 to 80 mM) and eluted with 0.5 M imidazole in buffer I. The eluted protein was dialyzed against buffer II (50 mM Tris pH 7.5, 100mM NaCl) and than against buffer II containing 40% glycerol. Protein was stored at –20°C in buffer II with 40% glycerol and 1mM DTT.

DszA, DszC oxidoreductase assays. *In vitro* desulfurization was carried out as described in Grey *et al*. (2) with slight modifications. Purified proteins (200pmols of DszA and 450 pmols of reductase) as well as crude lysates of *DszC* expressing cells were incubated with substrates in the presence of 10 µM FMN, 4 mM NADH, 100 mM NaCl, and 25 mM Na Phosphate buffer (pH 7.5). $DBTO_2$, 4-$MeDBTO_2$, BTO_2, 3-$MeBTO_2$, 5-$MeBTO_2$; 7-$MeBTO_2$ were used at 100µM concentration; 4,6-di$MeDBTO_2$ was used at 30µM concentration, and 2,4,8-tri$MeDBTO_2$ was used at 50uM concentration. DBTs and BTs were used at 100µM concentration. The reaction mixture was shaken at 230 rpm at 30°C. At designated time points the reaction was quenched with equal volume EtOH or acetonitrile. For the analysis of HPBS concentration quenched reaction mixture was then diluted 10X with 1% ammonium hydroxide. Substrate and product concentrations were determined by MS analysis.

Reductase assays. *In vitro* reductase assay was carried out as described in Grey *et al*. (2). The reductase activity was measured by the FMN-dependent oxidation of NADH monitored at 340 nm. Reaction was carried out at 30°C in 50 mM phosphate buffer, in the presence of 0.1 mM NADH and 20 µM FMN. Five to twenty pmols of enzyme was used.

3.0 Results and Discussion

Chemical synthesis (in support of multiple tasks)

We have continued to generate various benzothiophene (BT) and dibenzothiophene (DBT) precursors as model substrates and standards for analyses. This quarter, we have focused on generating a collection of mono-substituted methyldibenzothiophenes that will allow us to assess specificity of DszA and DszC enzymes. According to our analysis of the CED material, monosubstituted dibenzothiophenes are main contributors of sulfur among dibenzothiophenes. We also hope that activity against these model compounds will be indicative of activity against more heavily substituted dibenzothiophenes carrying substituents in the same positions as the model monomethyldibenzothiophenes.

We are using a converging approach described Tedjamulia *et al* (3). The dibenzothiophene core is assembled from appropriately substituted thiophenol and 2-bromocyclohexane precursors, followed by condensation and oxidative aromatization (Figure 1). We succeeded in preparation of 2-methyldibenzothiophene by this approach and directing our efforts towards completing of synthesis of 1-methyl and 3-methyldibenzothiophenes.

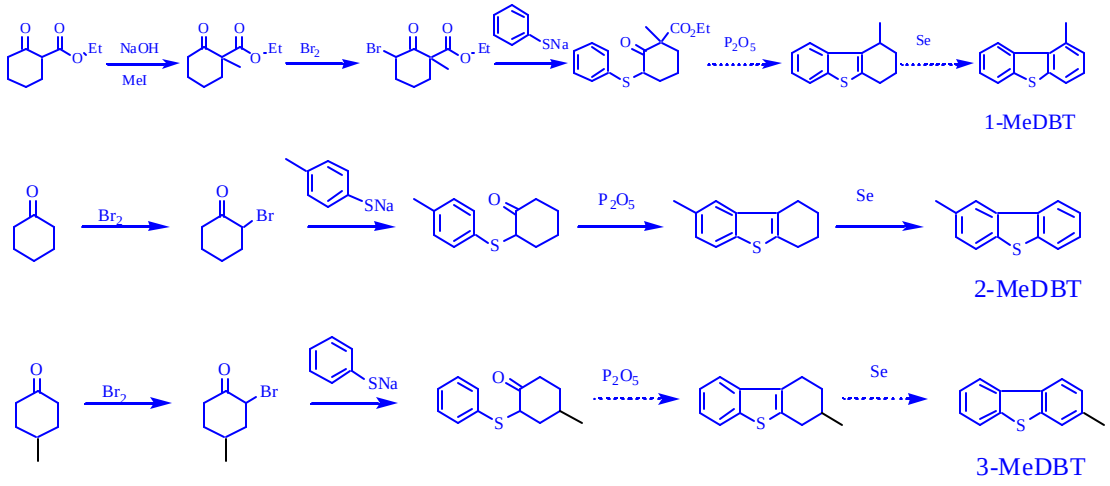


Figure 1. Schema for synthesis of 1-, 2-, and 3- substituted dibenzothiophenes.

Chemical Analysis (Tasks 1, 8 and support of other tasks)

Assay development. As previously described, we have chosen HPLC-MS-MS as a method of choice for determination of enzyme kinetics as well as the development of high-throughput procedures for enzyme evolution. Protocols for the analysis of potential BDS substrates have been developed and are described, as follows:

Substrate:	**Parent ion/fragment ion**
$DBTO_2$:	217.269/152.104
HBPS:	217.265/168.097

0.0	50%A	H2O/ACN gradient @1200 uL/min for 3 minutes
2.0	10%A	
2.1	50%A	
3.0	50%A	

Substrate:	**Parent ion/fragment ion**
4-MeDBBTO$_2$:	231.155/152.138
4,6-DiMeDBTO$_2$:	245.194/165.235
4-EtDBTO$_2$:	245.207/165.091
4,6-DiEtDBTO$_2$:	273.224/181.004
2,8-DiMeDBTO$_2$	245.057/165.255
H2O/ACN	25/75 isocratic mixture @ 1200 uL/min for 3 minutes
2,4,6 TriMeD BTO$_2$:	259.032/165.082
BTO$_2$:	167.127/103.149
2-MeBTO$_2$:	181.145/91.035
3-MeBTO$_2$:	181.108/91.051
5-MeBTO$_2$:	181.119/91.133
7-MeBTO$_2$:	181.152/91.025
H2O/ACN	20/80 isocratic mixture @ 1000 uL/min for 3 minutes
DBT:	184.085/152.107
DBTO:	201.116/183.971
4,6-DiMeDBT:	213.163/184.153
4-MeDBT:	198.137/165.170
ACN	100% @ 1000 uL/min for 3 minutes

We also developed and continue to develop protocols for the quantification of products of the DBT reactions as these will ultimately provide the highest degree of resolution of enzyme activities and a level of reproducibility required for high throughput assays to support directed evolution. Although relatively straightforward for DBT-derived products, BT-derived products can undergo a variety of spontaneous transformations, generating heterogeneous and potentially reactive products. With this in mind, we are

using BDS enzymes to produce products, in bulk, for determination of structure and reactive fate.

Molecular Biology and Biochemistry (Tasks 2, 4 and 9)

Cloning, Expression and Purification of Dsz Proteins

We have completed the cloning and expression of all available DszA-D proteins and each are in various stages of purification. Available enzymes are indicated in Table 1.

Enzyme	Organism	Native	N-His	C-His
DBT/BTO$_2$ monooxygenase				
DszA	***Rhodococcus* IGTS8**	+	+	+
DszA	***Nocardia* A3HI**	+	+	+
DszA	***Sphingomonas* AD109**	+	+	+
ToeA	***Tsukamarella* 670-1**	+	+	+
ToeA	***Tsukamarella* EMT4**	+	+	+
ToeA	***Nocardia* KGB1**	+	+	+
Desulfinase				
DszB	***Rhodococcus* IGTS8**	+	nd	nd
DszB	***Nocardia* A3HI**	+	nd	nd
DszB	***Sphingomonas* AD109**	+	nd	nd
DBT/BT monooxygenase				
DszC	***Rhodococcus* IGTS8**	+	+	+
DszC	***Nocardia* A3HI**	+	+	+
DszC	***Sphingomonas* AD109**	+	+	+
Flavin reductase				
DszD	***Rhodococcus* IGTS8**	+	+	+
Fre	***E. coli***	+	+	+
HpaC	***E. coli***	+	+	+

Table 1. *Cloned and expressed Dsz genes. nd=not done or planned currently*

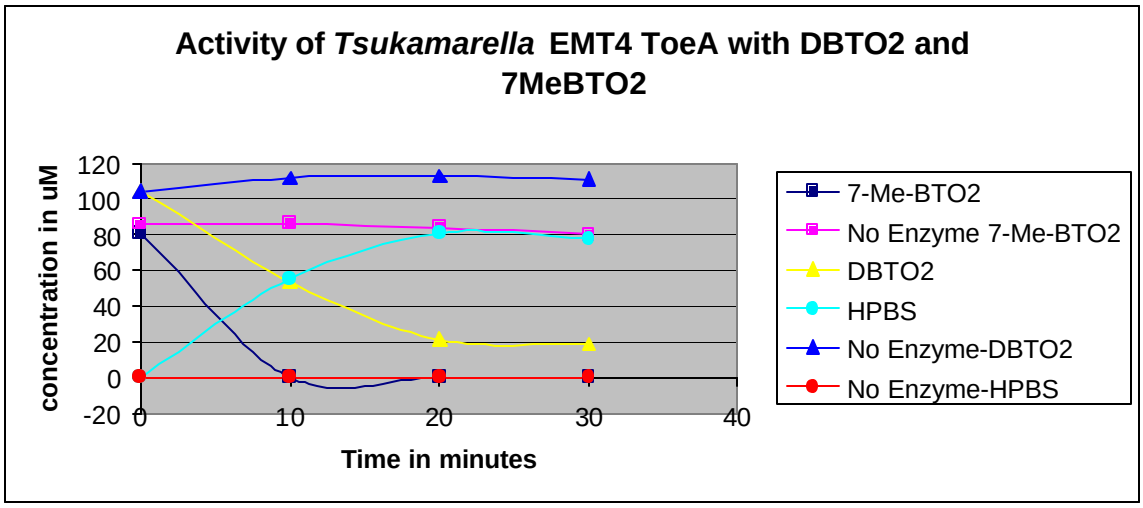


Figure 2. *Comparison of ToeA from Tsukamarella EMT4 (BD10799) activity with $DBTO_2$ and 7-$MeBTO_2$.*

Characterization of DszA activity

The effort described here is to determine the specific activity of different available DszAs with substrates relevant to biodesulfurization and, in light of revised goals, to a combined BDS/HDS process. With this in mind we focused on the following substrates: $DBTO_2$; 4-methyl$DBTO_2$, 2,4,6-dimethyl$DBTO_2$, 2,4,8-trimethyl$DBTO_2$, BTO_2, 3-methylBTO_2; 5-methylBTO_2 and 7-methylBTO_2. In the initial experiments we used crude lysates of *E. coli* expressing genes of interest with the addition of purified Fre protein. These initial experiments allow us chose the right conditions for the experiments to be performed with purified proteins. Since the majority of substrates are poorly soluble in water the specific activity is being determined at close to the saturation concentration. In these experiments we have used the rate of $DBTO_2$ oxidation to obtain relative rates. An example of the results of a typical kinetic reaction with lysate and substrates are presented in Figure 2.

A summary of rates so far obtained for DBT/BTO_2 monooxygenases is shown in Table 2. Although more experiments are planned, to be carried out with purified enzymes in order to obtain specific activities, some patterns can be observed. For instance, whereas the *Rhodococcus* and *Nocardia* DszA proteins demonstrate higher activity towards $DBTO_2$ and derivatives, the *Tsukamarella* ToeA proteins display a preference for BTO_2 compounds. Results obtained with *Sphingomonas* DszA are more ambiguous and experiments are in progress to resolve this. Even so it is clear that there are significant differences in the effects of the various substitutions and the relative rates as compared to unsubstituted $DBTO_2$ and BTO_2.

	DBTO2	*4-MeDBTO$_2$*	*4,6-diMeDBTO$_2$*	*2,4,8-trMeDBTO$_2$*	*BTO$_2$*	*3Me-BTO$_2$*	*5Me-BTO$_2$*	*7Me-BTO$_2$*
Rhodococcus IGTS8 DszA	1 (23.9+/-6.0)	0.4 (9.0+/-1.7)	0.2 (4.1+/-1.3)	0.5 (11.2+/-5.5)	0.16	0.12	0.3	0.3
Sphingomonas AD109 DszA	1	1.6	0.7	0.9	1.8	0.26	0.21	0.14
Nocardia A3H1DszA	1	0.1	0.7	2.6	0.01	0.03	0.03	0.9
Tsukamarella EMT4 ToeA	1	0.11	0.07	0.1	2.1	2.0	1.9	1.9
Tsukamarella 670-1	1	0.5	0.01	0.16	0.9	1.25	0.28	1.1
Gordonia KGB ToeA	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd

Table 2. *Activity of DszA and ToeA proteins with different benzothiophene dioxides.* The activity is expressed as the fraction of activity with $DBTO_2$ (expressed as 1). In cases the specific activity was determined with purified proteins, the values are given in brackets (in uM of substrate/min/nmol of enzyme). tbd = to be determined

DszC characterization.

We have also initiated enzymatic characterization of DszC and have so far performed reaction of crude lysates from *E. coli* cells expressing DszC. A kinetic reaction is shown in Figure 3.

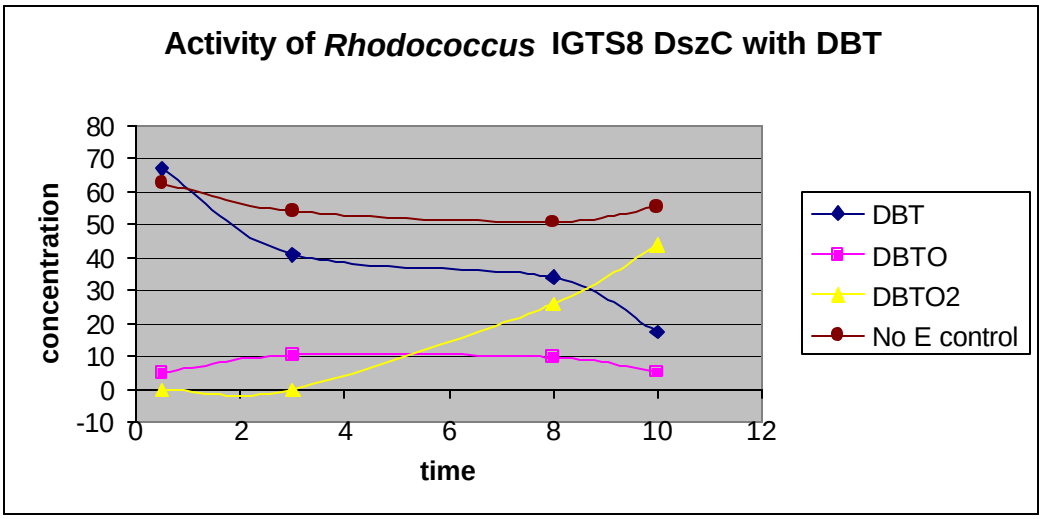


Figure 3. *Example of activity of crude lysate from E. coli cells expressing Rhodococcus IGTS8 DszC with DBT.*

	$DBTO_2$	4-$MeDBTO_2$	4,6-$diMeDBTO_2$	BTO_2	3-$MeBTO_2$	5-$MeBTO_2$	7-$MeBTO_2$
Rhodococcus IGTS8 DszC	++	+	+	-	+/-	-	-

Table 3. *Qualitative activity determination of Rhodococcus IGTS8 DszC with different dibenzo- and monobenzothiophenes.*

Fusion proteins.

We continued our investigation into the possibility of creating novel fusions between the IGTS8 DszA protein and one of two flavin reductases, DszD and Fre in order to determine whether the efficiency of the DszA reaction can be increased. With the purpose of determining the feasibility of this approach we constructed four different proteins in which DszA was coupled with flexible alanine-rich linker sequene to DszD or Fre – in one of two orientations – DszA-linker DszD (11902), DszA-linker-Fre (11903), DszD-linker-DszA (11904), and Fre-linker-DszA (11901).

All proteins were expressed and purified to determine both their respective reductase and oxygenase activities. As can be seen in Figure 4, all constructs except one (DszA-linker-Fre) had comparable activities to the Fre reductase itself. Unfortunately, under present assay conditions no oxygenase activities have been seen for any of the fusion proteins. We continue to investigate the properties of these fusion proteins.

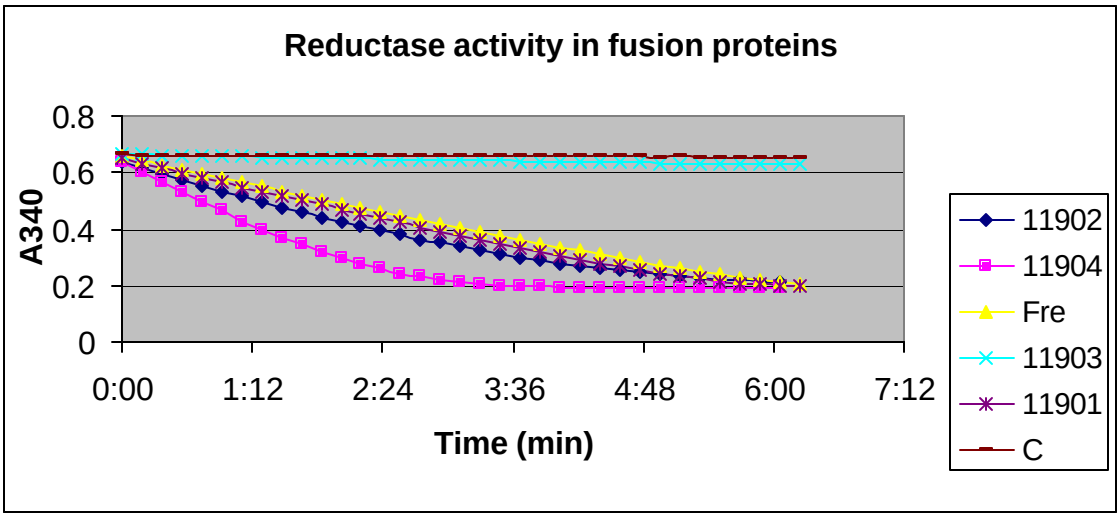


Figure 4. *Reductase activity of fusion proteins and comparison to Fre.* " C" = no enzyme control.

GSSM of *DszA* from *Rhodococcus* IGTS8.

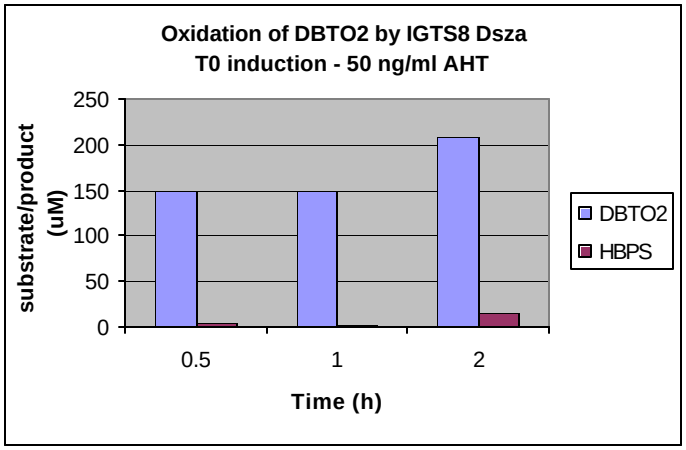
The generation and quality control of a GSSM mutant library of the *Rhodococcus* IGTS8 *dszA* gene has been completed. We are now determining optimal reaction and assay conditions for carrying an *in vitro* (as opposed to whole cell, where mass transfer of substrate into cells is an issue) high throughput characterization of this library with the goal to obtain variants with increased activity on substituted $DBTO_2$ and BTO_2 substrates.

We are currently designing the HTP assay to obtain variants with increased activity towards 4,6-diMe$DBTO_2$, as this compound (and its higher alkylated homologs) represents a significant component in Petro Star diesel and one that is problematic for hydrodesulfurization.

To effectively screen the GSSM library it has been necessary to carry out an extensive effort in determining parameters that are critical to reliability and interpretation of the results. To this end, multivariant analyses were performed using Design-Ease – a software package for determining critical parameters.

Results indicate growth and induction conditions are very important. Optimal conditions were determined to be induction at 30°C with 50 µg/ml AHT (to induce expression from recombinant plasmid). Time of induction was also determined to be significant and optimal conditions were obtained when the cells reach a density of OD_{600} = 1. Figure 5 shows an experiment where AHT inducer was added immediately upon inoculation and incubation vs. at a point where the culture reaches OD_{600} = 1.

Assay conditions were also determined be significant with influencing factors being the total concentration of DszA enzyme in the assay and the level of Fre reductase and aeration (as determined by degree of shaking). The amount of extraction reagent (B-PER is used to lyse cells) and temperature of incubation were not significant. Control of incubation time was also determined to be critical to insure coordination of the Fre reductase and DszA oxygenase activities.



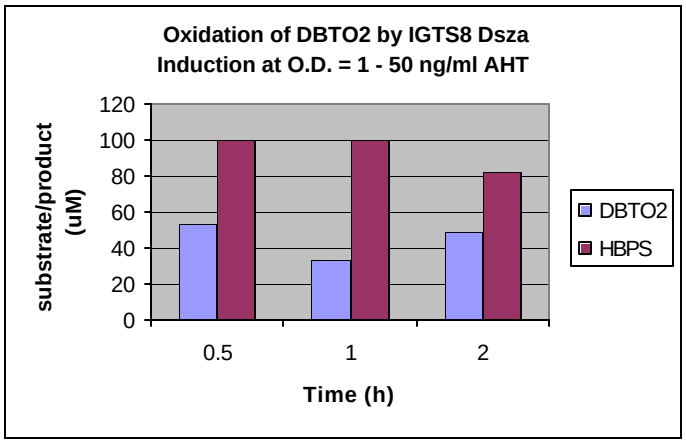


Figure 5. *Rates of $DBTO_2$ oxidation in relation to time of inoculation.*

A number of other parameters specific to implementing a robotic HTP assay have and continue to be investigated and these include:

1) Shaking vs. preaerated assay mixtures.
2) Lower temperatures to control reaction rates.
3) Clarified vs. crude cell lysates.
4) Timing of robotic maneuvers to insure uniformity from plate to plate and well to well.
5) Timing and stability of reagent additions (i.e. Fre reductase, FMN and NADH)

Development of optimal biocatalyst strain

Although *Rhodococcus* IGTS8 has been the prototype strain for biodesulfurization, it is likely other organisms may represent potentially better hosts for biocatalysis. Properties such as faster growth rates and general robustness may also be important properties in the development of an economically viable biocatalyst. With this in mind, we have begun the characterization of >130 *Rhodococcus* and related strains to identify potential hosts for the introduction of the biodesulfurization genes and comparison of properties to *Rhodococcus* IGTS8. *Rhodococcus erythropolis, R. globerulus, R. opacus, R. ruber* and *R. rhodochrous* strains, as well as *Gordonia ruberperfincta.* Strains have been obtained from both the ATCC and DSM culture collections. Phenotypic characterizations have included the ability to utilize rich and minimal media under a variety of aeration conditions and ease of plating to identify strains that may most easily be genetically manipulated.

AHT

From initial experiments it is clear a subset of strains have significantly more robust growth properties as compared to *Rhodococcus* IGTS8. In particular, a number of strains are able to grow to high density with significantly less aeration and may represent strains that are capable of utilizing available oxygen more effectively. This may be important in terms of both generation and utilization of the biocatalyst in the oxygen-requiring desulfurization reactions. Characterizations continue in an effort to identify a small subset of strains for introduction of the IGTS8 BDS plasmid and assessment of BDS activity (see next section).

Creation of IGTS8 BDS plasmid derivatives for introduction into *Rhodococcus* sp.

As mentioned above, a subset of *Rhodococcus* strains will be selected for introduction of the IGTS8 BDS plasmid and evaluation of biodesulfurization activity. Although the IGTS8 BDS plasmid is conjugal and can be mated into other *Rhodococcus* strains, it is necessary to generate antibiotic-resistant derivatives of each of the potential recipients to carry out the conjugation experiments. As this can be a very slow and inefficient process an alternative method is desirable.

We are currently exploring the use of transposome technologies to generate derivatives of the IGTS8 plasmid that can replicate and be mobilized from *E. coli* strains, thereby eliminating the need for generating special strains of each of the *Rhodococcus* strains chosen for characterization. Using an *in vitro* transposition reaction, we have successfully introduced an *E. coli* origin of replication (*oriR6K*) into the IGTS8 BDS plasmid and have demonstrated the stable maintenance of this plasmid in *E. coli* (Figure 6). By a further modification of the region transposed into the IGTS8 plasmid, to introduce an origin of conjugal transfer (*oriT*), we will be able to generate a derivative that can be efficiently introduced into *Rhodococcus* strains by conjugation from *E. coli.*

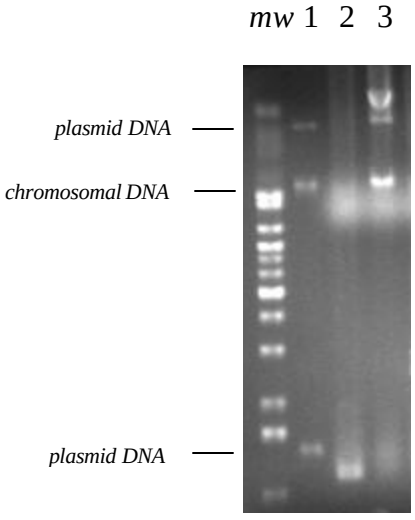


Figure 6. *Agarose gel of Rhodococcus plasmids replicating in E. coli.* Lane 1, Total plasmids purified form *Rhodococcus* IGTS8. Lane 2, small IGTS8 plasmid isolated from *E. coli* EC100 (*pir*); 3, Large IGTS8 plasmid replicating in *E. coli* EC100 (*pir*).

Nucleotide sequencing of the IGTS8 BDS plasmid

In order to advance our understanding of the BDS properties of IGTS8 as well as to provide information necessary for manipulation of the IGTS8 BDS plasmid we have begun an effort to sequence the plasmid complement of *Rhodococcus* IGTS8. Studies in the literature have suggested there are at least two plasmids in this strain (5) but nucleotide sequencing will be required to definitively answer this question.

Total plasmids have been prepared from IGTS8 using a modified alkaline lysis procedure. From initial agarose gel analysis it appears there is at least one high molecular weight plasmid as well as a very low molecular weight plasmid of between 2 and 10kb. To generate molecules for sequencing a *Sau*3A-partial recombinant library has been constructed and the sequencing has begun. Early analysis confirms the presence of at least two plasmids although more extensive sequencing and assembly will be required to resolve this observation.

Development of screens and selections for BDS activities.

Work continues in areas previously described to develop screens and selections for BDS activity that can be used for both gene discovery and evolution of improved activities. Two alternative approaches: a sulfur selection strategy and a bioreporter strategy were previously described and development continues as described below. In addition, a third approach, the development of a fluorescent BDS product approach, was also initiated and will be described.

E. coli strains for sulfur selection.

We have completed the generation of an *E. coli* strain deleted for the *cysP* operon and have demonstrated that it is unable to utilize sulfate both in liquid or solid media. In addition its ability to utilize the residual sulfur present in solid agar or agarose media is almost completely eliminated as compared to the wild-type strain. It is still able to grow as well as wild-type on rich media or minimal media containing sulfite or thiosulfate as sole source of sulfur. This latter observation is surprising, as it has been reported the *cysP* operon encodes a thiosulfate transporter and its elimination should abolish the ability to utilize thiosulfate (4). This result remains to be resolved but does not directly affect the use of this strain for sulfur selection in a BDS screen.

The *dszB* genes from *Sphingomonas* AD109 and A3HI have been cloned and introduced into both the *cysP* mutant as well as the wild-type strain. The expression of these genes confers the ability to utilize hydroxybiphenylsulfinate as sole source of sulfur, although dependent on the appropriate level of AHT addition (data not shown). There was, however, a significant delay in the growth of the *cysP* mutant as compared to

the wild-type strain that is not readily explained. This is also apparent when comparing the utilization of HBPS on solid medium, with growth of the mutant being much slower and less vigorous than wild-type. This suggests that the deletion of the *cysP* operon has some additional unexplained effect on HBPS or sulfite utilization that remains to be resolved.

Bioreporter plasmid

We previously described the construction of a reporter plasmid able to generate both a fluorescent signal in response to hydroxybiphenyl. In this quarter we constructed a derivative of this reporter that allows also for the expression of an antibiotic resistance (chloramphenicol) and is titratable to the level of HBP present. A number of preliminary experiments have been carried out to determine the utility of the construction for use in discovery and evolution of the DszB enzyme as well as the BDS pathway as a whole. Unfortunately, control experiments indicated at least two problems with the system.

Introduction of a DszB-expressing clone in an *E. coli* strain with the above described reporter did not produce detectable signal in liquid or solid media containing hydroxybiphenylsulfinate as substrate for DszB. It is apparent the levels of free HBP that are produced are insufficient to induce expression of the reporter under these circumstances. This may occur because of the relatively large volume of culture medium in relation to a single DszB-expressing cell coupled with the diffusion of HBP out of the cell. We are currently exploring the possibility of using our Gigamatrix technology in which culture volumes are in the nanoliter range and may address the diffusion/concentration issue.

The second issue that has arisen with the use of the reporter plasmid is the frequency of false positives that are obtained when *E. coli* cells are plated in the absence of inducer compounds and selected for resistance to chloramphenicol. These apparent mutants occur with varying frequencies (from $10e^{-5}$ to $10e^{-7}$), depending on the nature of the reporter construct. Even the reporter plasmid by itself, without the *hbpRC* sequences that respond to HBP, gives a detectable level of false positives. To determine the cause of these false positives, a number of chloramphenicol-resistant clones from the reporter plasmid alone (without the *hbpRC* sequences) were selected and the nucleotide sequence of the reporter plasmid was determined. As can be seen in Figure 7**,** IS sequences (insertion elements) of both the IS3 and IS10 family have appeared to transpose upstream of the chloramphenicol acetyltransferase gene, within or immediately downstream of the resident terminator sequences. Insertion elements are found scattered throughout the sequence of most bacterial strains and can transpose into multiple DNA sites, including foreign plasmids. Many IS sequences contain active promoters and can induce expression of genes that normally would not be expressed. This observation suggests the low number of false positives seen with the bioreporter may not be avoidable and must be taken into consideration in any discovery or evolution platform in which it is to be used.

Experiments continue to determine the extent of this limitation and develop solutions that may still allow this strategy to be used.

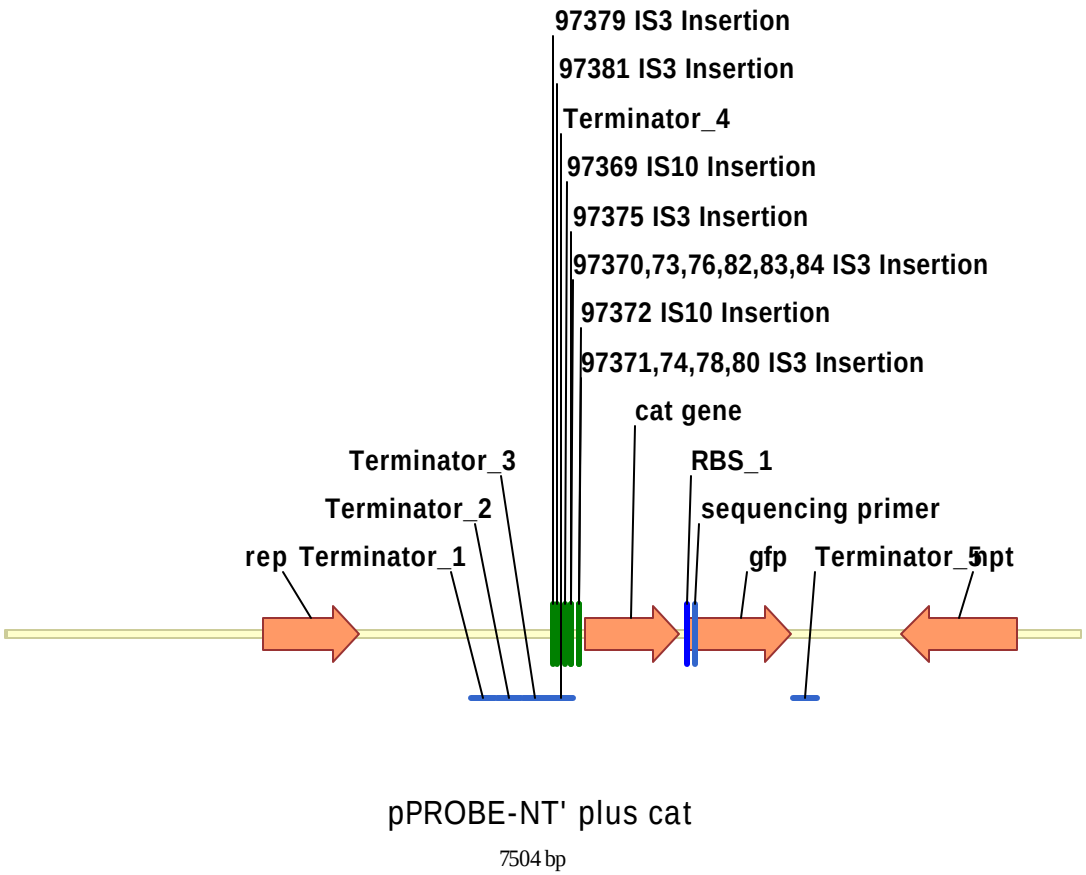


Figure 7. *Schematic of bioreporter plasmid showing sites of IS insertions that lead to false-positive expression of cat gene (encoding chloramphenicol resistance).* gfp = Diversa green fluorescent protein

Hydroxybiphenyl fluorescence

Although this strategy is not itself novel, we are exploring the development of using the fluorimetric detection of 2-hydroxybiphenyl (HBP) as based on the cell free method of Gray et al (2). To determine whether the HBP assay could be effectively used in the presence of live cells a number of preliminary experiments were performed to determine the assay feasibility. HBP of varying concentrations was diluted in a variety of solutes: water, Luria-Bertani (LB) medium (1) and a Diversa-developed low fluorescence medium, CM6, for use in HTP applications such as Gigamatrix. In addition, HBP was diluted in a mixture of *Esherichia coli* cells grown in CM6. Aliquots of these mixtures were then measure for fluorescence, with excitation at 294 nm and measuring the emission at 420 nm (Figure 8-10).

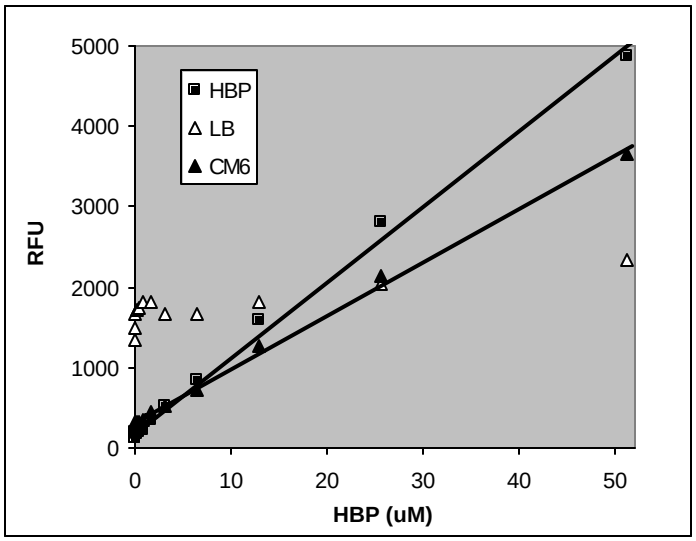


Figure 8. *Standard curve of relative fluorescent units (RFU) versus concentration of HBP (mM).* The emission of HBP at 414 nm (ex. 288 nm) was measured after dilution in water, LB or CM6 media.

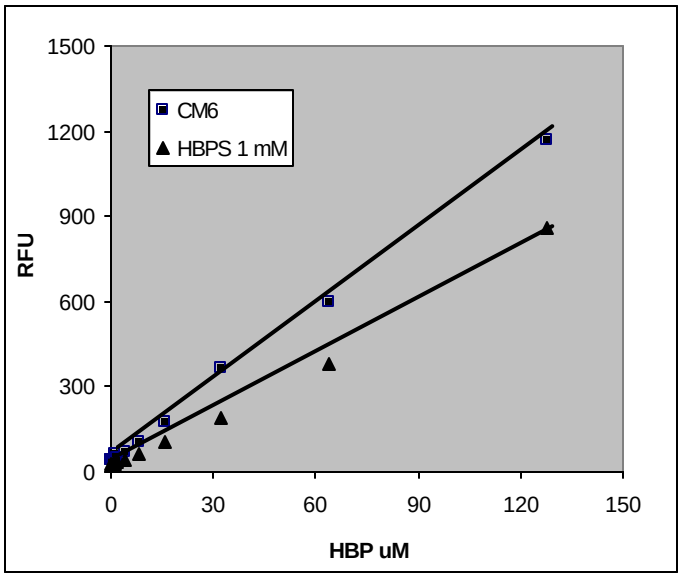


Figure 9. *Standard curve of relative fluorescent units (RFU) versus concentration of HBP (mM)*. The emission of HBP at 420 nm (ex. 294 nm) was measured after dilution in working strength CM6 medium or HBPS (1 mM).

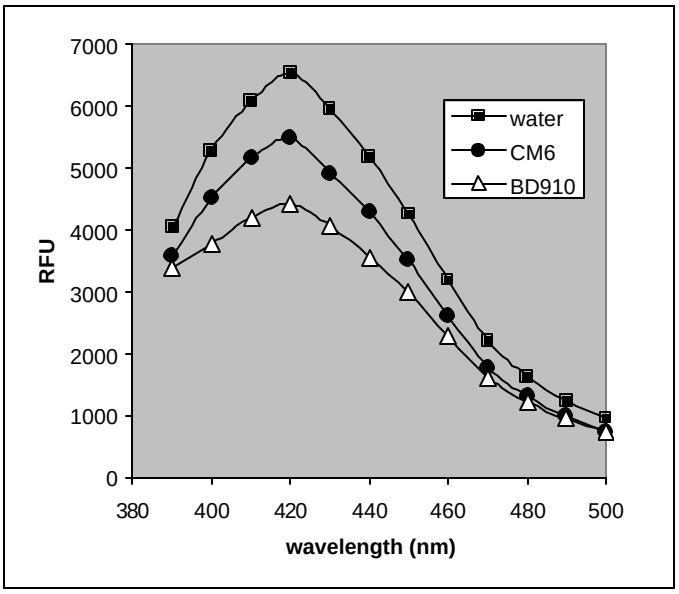


Figure 10. *Emission spectrum of HBP (51.2 mM) diluted either in water, working strength medium (CM6) or CM6 containing 4 OD_{600} E. coli cells (BD910).* Excitation was fixed at 294 nm and emission (RFU) was monitored between 390 and 500 nm.

As can be seen in Figure 8, although HBP can be detected in water careful consideration of the medium composition is required. LB interferes with the detection of HBP whereas CM6 does not. In addition, a high level of HBPS (which would be necessary in an assay for DszB activity, for example) does not quench the signal generated by HBP (Figure 9). Finally, a high concentration of cell mass does not significantly interfere with HBP detection (Figure 10).

We are currently carrying out pilot *in vivo* experiments using cloned *dszB* genes expressed in *E. coli* to determine optimal conditions for use in both a1536-well assay format as well as the 100,000 well GigaMatrix plate system. Once refined, we will explore the use of this method as a complementary approach for the discovery and evolution of *dszB* genes.

4.0 Conclusions

The chemical analysis of PetroStar diesel and CED extract has been completed and a report detailing the results is attached. We are currently compiling a new task list for the new direction of our research. Upon completion and submittal we will await the authorization to begin characterization of partially hydrodesulfurized diesel fuel

In the mean time, we completed the cloning and expression of all available biodesulfurization genes and have carried out a significant amount of work determining enzymatic properties in regard to rates and specificities. Development of *in vitro*

analytical techniques to support the evolution and optimization of biodesulfurization enzymes has been carried out and continues.

We completed the generation of a *Rhodococcus* IGTS8 *dszA* GSSM mutant library and are developing the protocols for a robust and reliable high throughput screen to improve the activity of DszA on substrates determined to be most relevant for a combined BDS/HDS process (see attached Task Revision document). We initiated an extensive characterization of alternative hosts for biocatalyst genes and are genetically modifying the *Rhodococcus* IGTS8 plasmid from its introduction into strains identified as having the most desirable properties.

Finally, we initiated the complete sequence determination of the *Rhodococcus* IGTS8 plasmid complement in order to gain a greater understanding of the genes required for biodesulfurization by this strain and to assist in the genetic manipulation of the BDS plasmid.

5.0 References

1. Sambrook, J., E. F. Fritsch, and T. Maniatis. (1989) Molecular cloning: A laboratory manual, 2nd Ed. Cold Spring Harbor Laboratory Press.

2. Gray K.A., O.S. Pogrebinsky, G.T. Mrachko, L. Xi , D.J. Monticello and C.H. Squires. (1996) *Nat Biotechnol*. **14**:1705-1709.

3. Tedjamulia, M. L., Y. Tominaga,, R.N. Castle. and M.L. Lee, M. L. (1983) *J. Heterocyclic Chem.* **20**:1485-1496.

4. Sirko, A., M. Zatyka, E. Sadowy and D. Hulanicka (1995**)** *J. Bacteriol*. **177**:4134-4146.

5. Denis-Larose, C., D. Labbe, H. Bergeron, A.M. Jones, C.W. Greer, J. al-Hawari, M.J. Grossman, B.M. Sankey and P.C. Lau (1997**)** *Appl. Environ. Microbiol*. **63**:2915-9.

6.0 List of Acronyms and Abbreviations –

BDS	Biodesulfurization
CED	Conversion Extraction Desulfurization
DBT	Dibenzothiophene
$DBTO_2$	Dibenzothiophene sulfone
DNA	Deoxyribonucleic acid
$FMNH_2$	Reduced flavin mononucleotide
GC	Gas chromatography
HBP	Hydroxybiphenyl
HBPS	Hydroxybiphenylsulfinate

HPLC	High performance (pressure) liquid chromatography
MS	Mass spectrometry
NADH	Reduced nicotinamide dinucleotide
PCR	Polymerase chain reaction
RP	Reverse phase
SCD	Sulfur chemiluminescence detector

Analysis of Sulfur Compounds in Petrostar Chemical Extraction Desulfurization Materials

Michael D. Levin, Ph.D.
Integrated Chemical Processes Department
Diversa Corporation

DIVERSA

TABLE OF CONTENTS

Introduction

This report describes the analytical approach used to determine the composition of sulfur-containing compounds in Petrostar chemical extraction desulfurization (CED) material. The origin and nature of the samples, their handling, methodology of analysis and detailed procedures are explained. Qualitative and quantitative results of analysis are discussed.

Background

Petrostar marine diesel is obtained by straight run distillation from North Slope Alaskan crude oil. It is not subjected to cracking, reforming, hydrotreatment, or any other condition, which might change the chemical composition of substances present in the crude. Only the boiling point range of the diesel fraction and the crude composition limit the diversity of structures. Thus, the number of sulfur-containing compounds present in detectable concentrations is very high. Attempts to analyze the composition of these compounds by individual methods of separation yield only very broad unresolved humps with few structural features. The retention properties of sulfur-containing organic compounds are very similar to each other, and to those of hydrocarbons, which are present in a much higher concentration in diesel. This makes speciation even more problematic.

Sulfur-containing compounds found in diesel fractions may be divided into two large groups: aromatic thiophenes and aliphatic sulfides. In aromatic thiophenes, sulfur is part of an aromatic ring structure, while in aliphatic sulfides sulfur atoms do not participate in aromatic conjugation. Aromatic thiophenes may be divided further into thiophenes (Thios), benzothiophenes (BTs), and dibenzothiophenes (DBTs). In aliphatic sulfides the sulfur atom may be attached to saturated hydrocarbon substituents, cyclic or aromatic substituents, or be part of a cyclic structure itself. The presence of thiols and disulfides in diesel fraction distillates has been reported as well.

Petrostar marine diesel contains around 3500 ppm total sulfur. The distillation range of the diesel fraction defines the range of molecular weights for each class of compounds. For all classes but DBTs, this range corresponds to heavily alkylated core structures. The number of possible isomers grows exponentially with the degree of alkylation, and so does the complexity of mixtures.

One of the commonly used techniques to simplify the separation of sulfides and thiophene derivatives is initial oxidation to the corresponding sulfones and analysis of the resulting mixtures. With the notable exception of Thios, oxidation of sulfides to sulfones can be done selectively and quantitatively. Oxidation usually produces one well-defined product, which can be easily traced back to the original sulfide. Sulfones can be easily separated from the matrix of hydrocarbons since they are more polar than sulfides. This separation may be achieved by selective extraction or by chromatography. Sulfones also differ in retention properties among themselves much more than the original sulfides.

Oxidation of thiophenes occurs at a lower rate than oxidation of benzothiophenes, dibenzothiophenes, and aliphatic sulfides. The reaction results in a complex mixture of products of side reactions, the most prominent of which is Diels-Alder addition. Thiophene sulfoxides and thiophene sulfones are strong dieneophiles, and can react with themselves, or with starting thiophenes and benzothiophenes. Extensive alkylation of the thiophene framework reduces the Diels-Alder reactivity of the corresponding sulfoxides and sulfones, rendering them somewhat more stable.

Experimental

CED material

The CED material was obtained by chemical oxidation of Petrostar stock diesel with peracetic acid and extraction of polar products with acetic acid. A sample of CED material was provided by Petrostar and analyzed according to a procedure outlined in a flowchart in **Figure 1**.

Sample manipulations

A sample of CED material was separated into fractions by preparative HPLC on a C_{18} reverse phase column. Fifty fractions of equal volume were collected. Solvents were evaporated from each of the fractions. An *i*-propanol solution of two internal standards, butyl sulfone and pentyl sulfone, was used to redissolve residues from fractions. The reconstituted fractions were analyzed by gas chromatography-

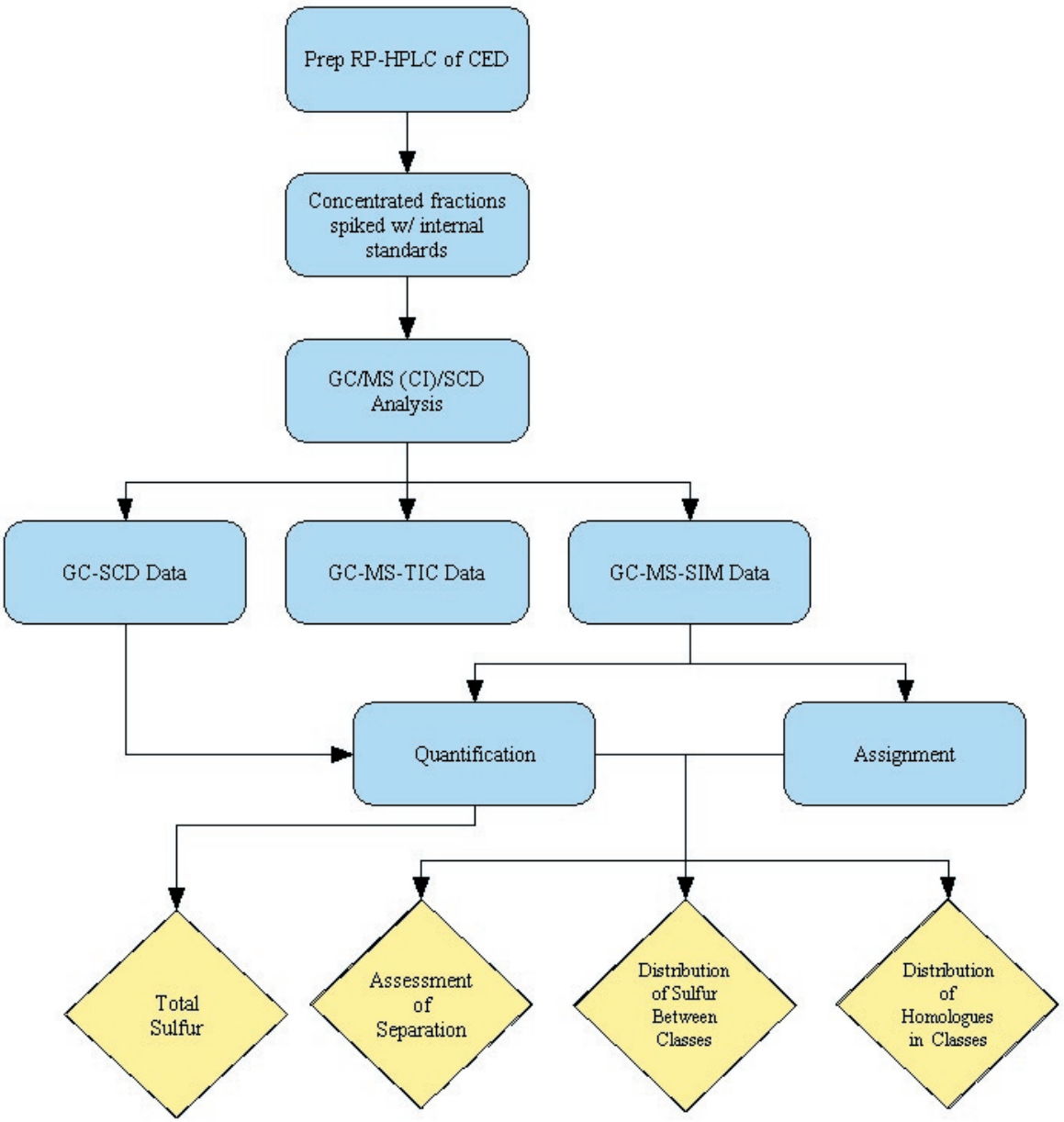


Figure 1. Flow-chart describing the analytical techniques used to analyze and characterize the Petrostar CED material.

mass spectrometry (chemical ionization)-sulfur chemiluminescence detection (GC/MS(CI)/SCD). GC inlet split ratios were optimized for each fraction such that the maximum signal before saturation was obtained in the SCD detector. Two injections were done for each sample. The first injection was used to obtain an estimate for the optimal split ratio and to acquire a mass spectrometer total scan total ion chromatogram (TIC). Ions observed in the TIC mode were used to perform selected ion monitoring (SIM) during the second run with the optimal split ratio. Only the data of the second set of chromatograms were used in analysis.

Data analysis

The sulfur-containing species from the CED material eluted between 13 and 21 minutes under the GC conditions. The SCD trace was integrated between these times, and the integral was converted into molar concentration of sulfur by comparison with the integrals of two internal standards. The integrated intensities of individual ions from SIM data were used to estimate concentrations of isomeric species in each fraction. These calculations were done under assumptions of (i) equal ionization efficiency for different species and (ii) no significant ion suppression in cases when compounds were not separated before they entered an ionization chamber of the mass spectrometer.

Five classes of compounds were identified in the CED material fractions based on the mass spectrometric data: benzothiophene sulfones (BTO_2s), dibenzothiophene sulfones ($DBTO_2$s), thiophene sulfones ($ThioO_2$s), bicyclic aliphatic sulfones ($BicycO_2$s), and monocyclic aliphatic sulfones ($MonoO_2$s). Compounds within these classes were identified as members of a homologous series differing from one another by an integral number of CH_2 fragments. Masses of observed [M+H] ions for different homologues are listed in **Table 1**. Concentrations of various homologues belonging to the same class of compounds were summed up across fractions and provided the total concentration of compounds belonging to this class. SCD data from different fractions were normalized for different split ratios.

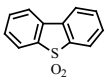
Classes	Base structure	[M+H] for Cx homologues
$DBTO_2$		217 + 14x
BTO_2		167 + 14x
$ThioO_2$		117 + 14x
$MonoO_2$		121 + 14x
$BicycO_2$		189 + 14x

Table 1. Masses of observed [M+H] ions for different homologues.

Results and Discussion

Separation

HPLC-UV traces of preparative separations (not shown) are not informative due to overloading of the detector and a lack of resolution on this step. GC/SCD/MS traces of individual fractions show various degrees of separation. The first four fractions (eluting at 11 – 12.5 min. on HPLC) show good separation of peaks down to baseline. In the next seven fractions (13 – 16 min.) many peaks overlap to some degree. Starting with the fraction eluting at 16.5 min. on HPLC, the GC chromatograms develop an unresolved hump between 14 and 20 minutes with peaks on top of it. The contribution of the hump steadily increases from fraction to fraction relative to the contribution of resolved peaks. The retention time of the signals increases as well. The last nine chromatograms (fractions 32 – 35.5) display mostly unresolved signal between 17 and 21 minutes. Every gas chromatogram contains peaks for two internal standards: butyl sulfone (retention time 8.3 min) and pentyl sulfone (retention time 11.5 min).

Integration of GC/SCD traces normalized to internal standards revealed the total sulfur distribution between fractions (**Figure 2**). The graph shows four peaks and three valleys between them. The first peak (fraction 12.5) is attributed to elution of C1 $DBTO_2$s. C2 $DBTO_2$s and C5 BTO_2s form the second peak (fractions 14.5 and 15). The peak at fraction 17 is due to elution of C6 BTO_2s, the most abundant homologue among alkylated BTO_2s. The last broad peak around fraction 19.5 is composed of a variety of compound classes with some dominance of C7 BTO_2s. Total sulfur content in the initial CED material is 15.4 g/l.

A contour plot of SIM data (**Figure 3**) allows the assessment of the separation achieved by 2D chromatography. Separation must be assessed along several lines: (i) separation among classes of compounds, (ii) separation among series of homologues within classes, and (iii) separation among individual compounds within series of homologues.

(i) The contour plot of SIM data reveals that $DBTO_2$s were nicely separated from all other classes of compounds present in CED. $ThioO_2$s, on the other hand, overlap extensively with BTO_2s and $BicycO_2$s.

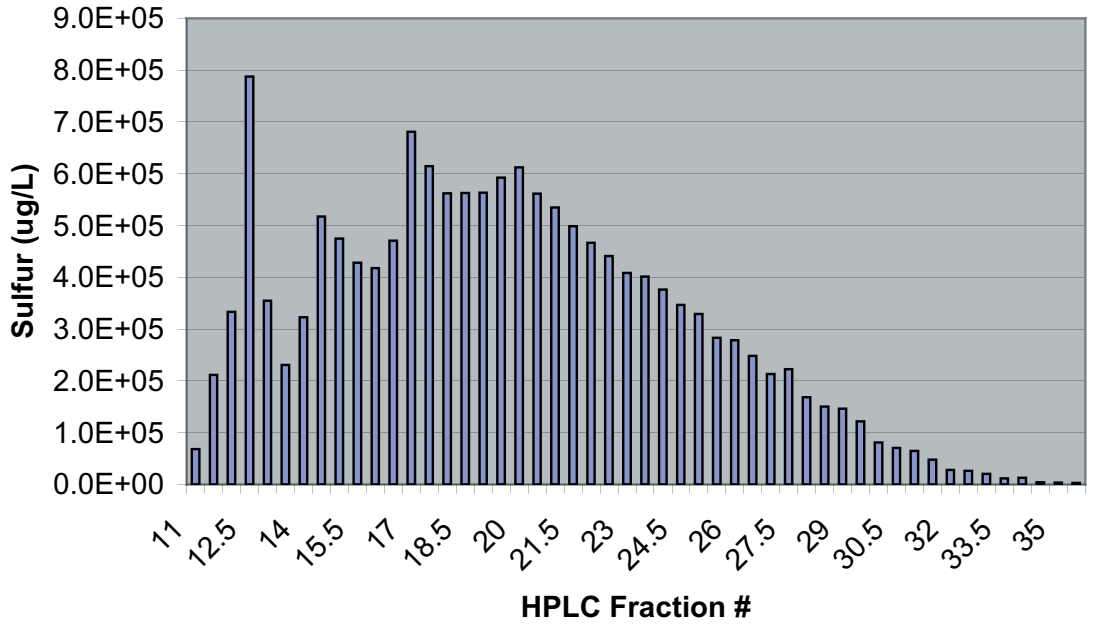


Figure 2. Total sulfur distribution between fractions.

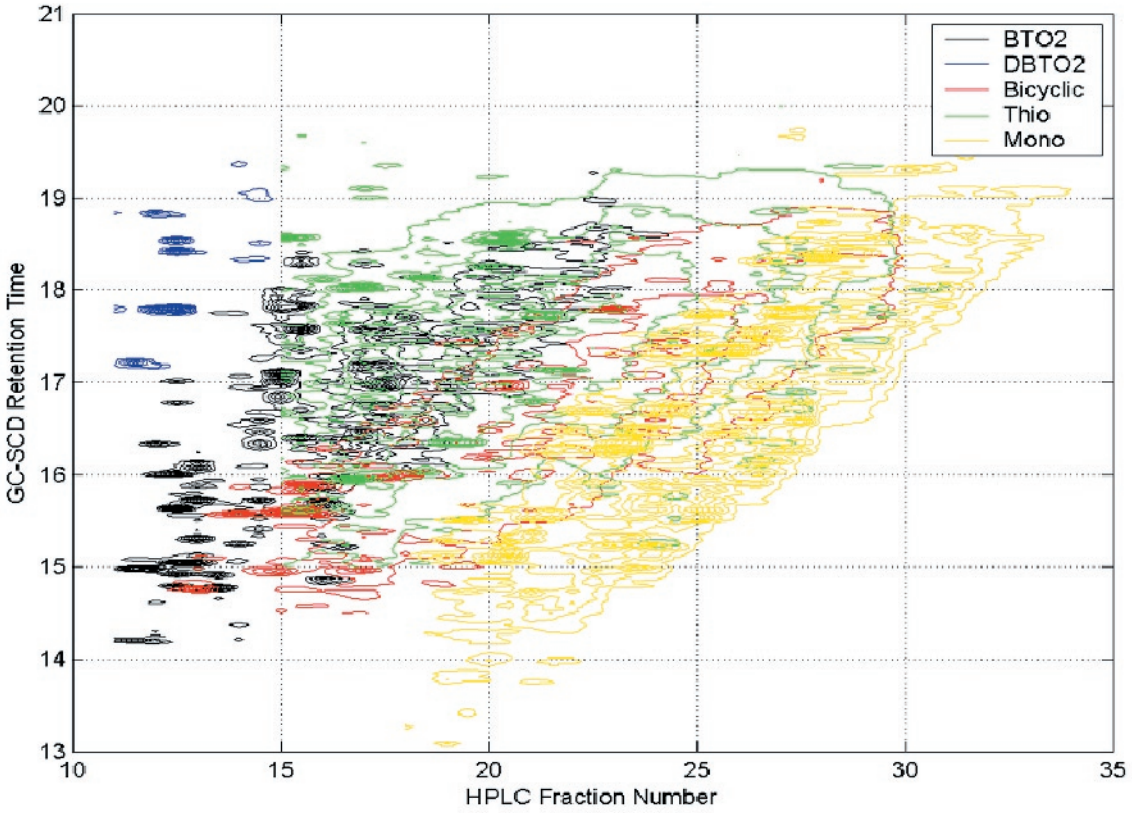
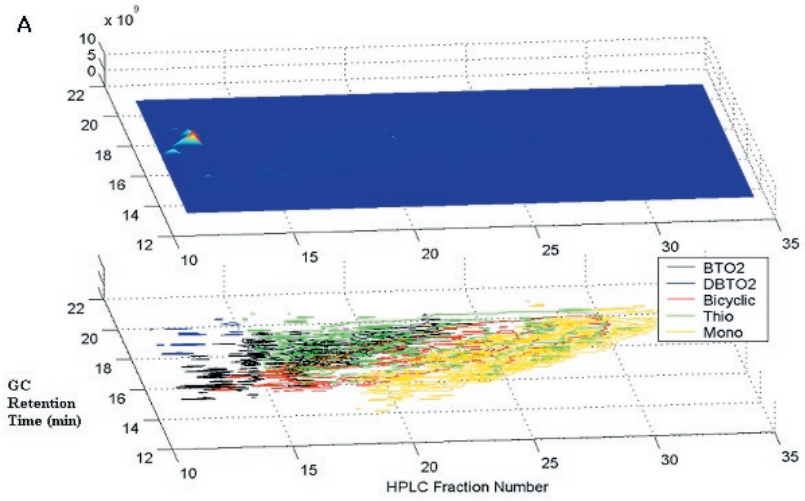
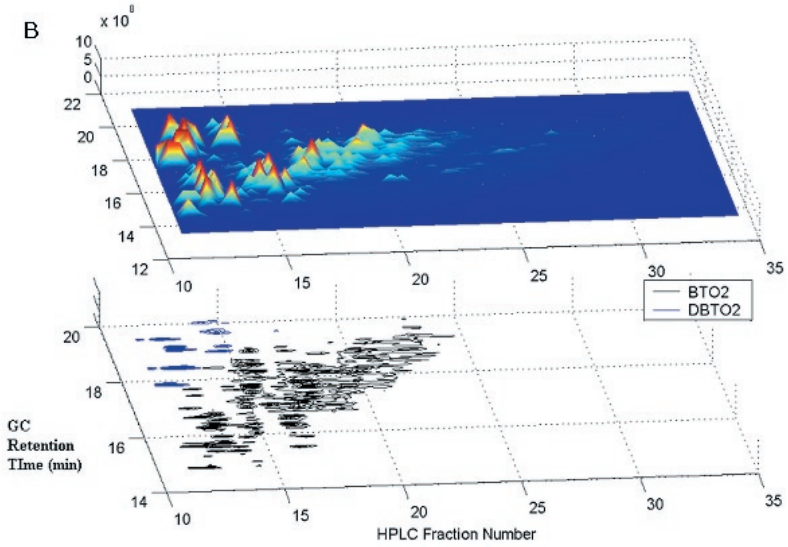


Figure 3. A contour plot of SIM data.

BTO_2s and $BicycO_2$s have some overlaps, especially in the area between HPLC fractions 12.5 and 16.5 and GC retention times between 14.7 and 16 minutes. Quite good separation was achieved for $MonoO_2$s. Contours of $ThioO_2$s and $BicycO_2$s overlapping with the $MonoO_2$ area represent only a low concentration of the contaminating species.

The intensities of the peaks corresponding to different classes of compounds are shown on surface plots of SCD data (**Figure 4**). The landscape is clearly dominated by a single peak of 4-methyldibenzothiophene sulfone (**Figure 4a)**. All $DBTO_2$s combined contribute a minor fraction to the overall sulfur content of the CED material, but the relatively small number of isomers of $DBTO_2$ in the mixture provide for high intensity of peaks corresponding to individual compounds. In order to observe peaks corresponding to other compounds, the SCD count vertical scale has to be expanded. Peaks of BTO_2s (**Figure 4b**) cover an underlying plateau of $ThioO_2$s. The intensity of individual peaks steadily decreases along with the HPLC fraction numbers and the number of possible isomers corresponding to the homologues. Some smaller peaks on the front edge of the BTO_2 ridge can be assigned to $BicycO_2$s. These peaks are more visible in **Figure 4c**, along with the $MonoO_2$ ridge in front and to the right of them.





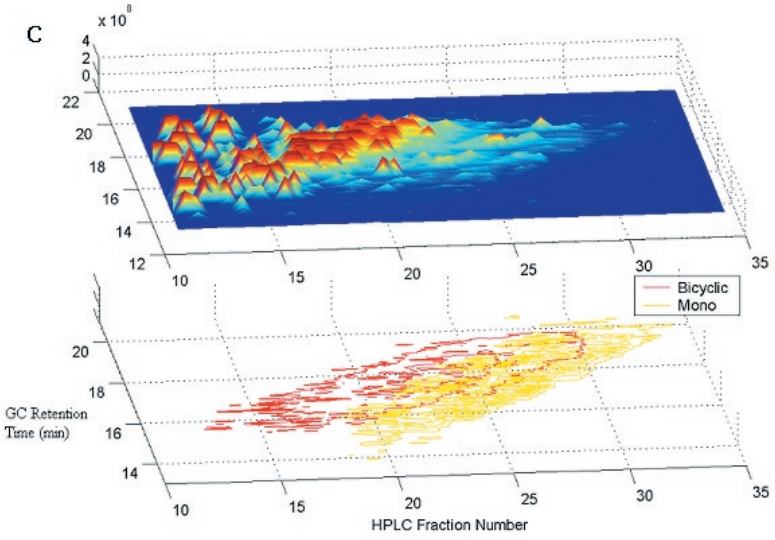


Figure 4. Surface plot of SCD intensities and contour plot of SIM intensities for 2-D chromatography of CED material

a) Full vertical scale for SCD and contours for all five classes of compounds

b) Expanded 10x vertical scale for SCD and contours for $DBTO_2$s and BTO_2s

c) Expanded 25x vertical scale for SCD and contours for $BicycO_2$s and $MonoO_2$s

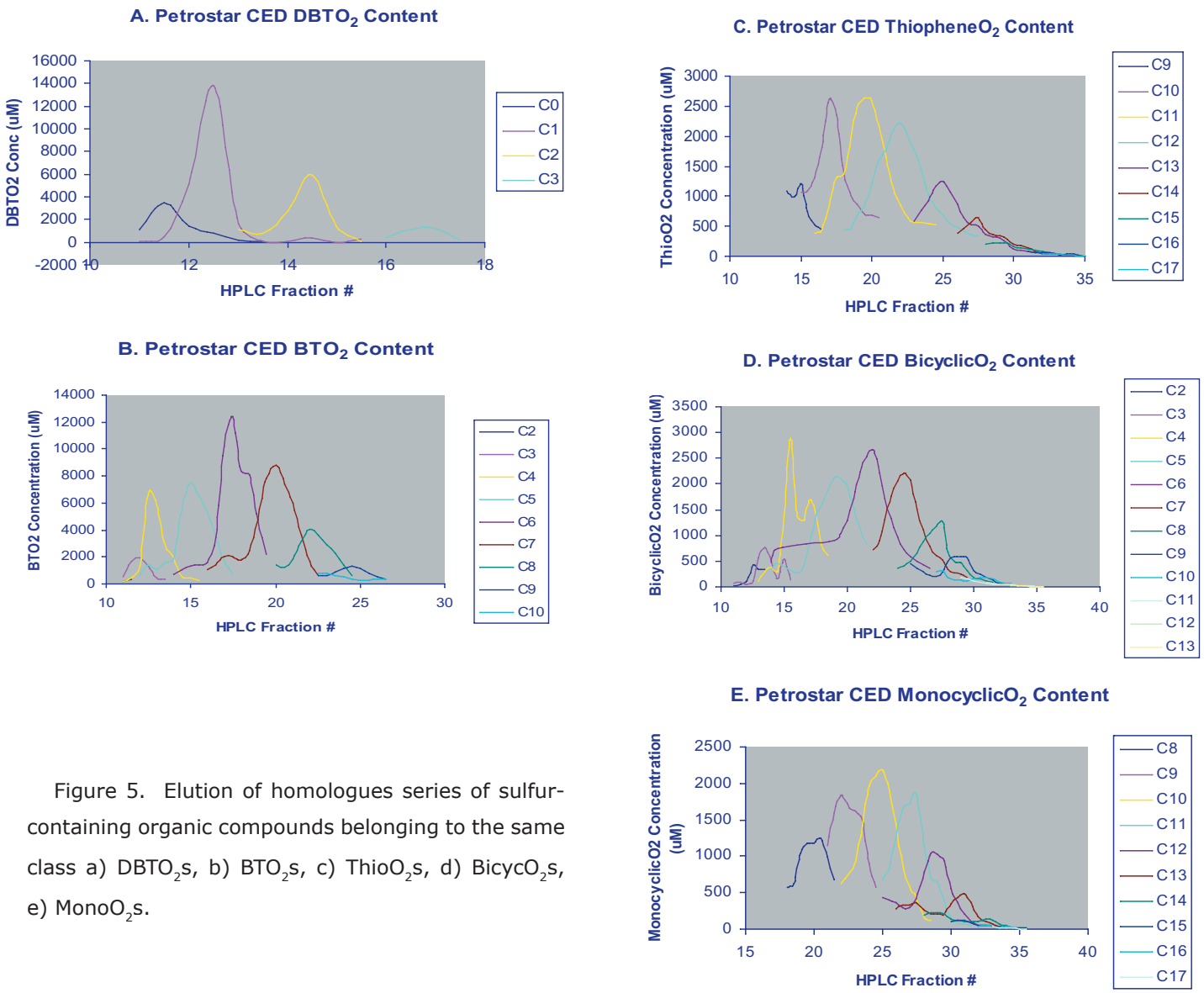


Figure 5. Elution of homologues series of sulfur-containing organic compounds belonging to the same class a) $DBTO_2$s, b) BTO_2s, c) $ThioO_2$s, d) $BicycO_2$s, e) $MonoO_2$s.

(ii) Separation within series of homologues within classes is achieved primarily by HPLC (**Figure 5**). GC improves separation only when the number of isomers for homologues is small and individual compounds can be separated (fractions 11 – 12.5). In the latter fractions, signals corresponding to different homologues overlap significantly. Distributions of sulfur among homologues within classes of compounds have broad maxima characteristic for each class (C1 for $DBTO_2$s (**Figure 5a**), C6 for BTO_2s and $BicycO_2$s (**Figure 5b**), C11 for $ThioO_2$s (**Figure 5c**), and C10 for $MonoO_2$s (**Figure 5d**). The boiling range of the diesel fraction probably defines these maxima. All distributions but the one for $DBTO_2$ have bimodal character. This effect is most prominent for C4 $BicycO_2$s (**Figure 5d**). This bimodality may be explained by the presence of two prevailing trends in substitution patterns of the core structures with different elution properties, each of which has its own distribution pattern. More experiments will be needed in order to study this phenomenon in details.

(iii) GC separation of individual compounds is achieved only in the first four HPLC fractions, containing a limited number of $DBTO_2$s and BTO_2s. In the rest of the fractions, the presence of an overwhelming number of individual compounds precludes their separation.

Assessment of accuracy of various aspects of analysis

The experiments described here were designed to obtain the most accurate analysis possible. However, we worked under several assumptions, and we feel obliged to discuss the possible effects that these assumptions may have on the accuracy of our analysis. CED material was obtained by oxidation of a sample of Petrostar marine diesel and the subsequent extraction of polar compounds with acetic acid. We assumed that (i) oxidation quantitatively converts all sulfides and thiophenes into the corresponding sulfones and that (ii) acetic acid completely and selectively extracts all sulfones out of the matrix of diesel hydrocarbons. Assumption (i) is not true for substituted thiophenes. Thus, the concentration of their oxidation products can be significantly lower in the CED material compared to the concentration of thiophenes in diesel. In addition, acetic acid may be less efficient in extraction of saturated aliphatic sulfones due to their lower polarity. Thus, the fact that we did not find dialkyl sulfones in the CED material may be attributed to an incomplete extraction.

The SCD data is quantitative and does not require any assumptions to yield the values for total sulfur content of individual fractions and of the whole CED material. However, TIC and SIM data were processed under two major assumptions: (i) ionization efficiency of individual compounds is the same, and (ii) compounds do not compete for ionization when they enter the mass spectrometer as a mixture. In order to remove the dependence of our analysis on the first assumption we would have to synthesize all individual compounds and measure standard curves for them under the ionization conditions used. This is a formidable task. The second assumption was validated by an extensive separation achieved by the 2D chromatography and by the fact that the amount of overlapping samples entering the ionization chamber of the mass spectrometer did not exceed its overall ionization capacity. Still, ThioO_2s were not separated from BTO_2s and BicycO_2s, and our estimate of their contributions may not be completely accurate.

Conclusions

Sulfur-containing organic compounds of the CED material of Petrostar marine diesel was separated by 2D chromatography. The CED material contains 15.3 g/l total sulfur, distributed between five classes of sulfur-containing organic compounds. BTO_2s have most of the sulfur (43%) with BicycO_2s, ThioO_2s, and MonoO_2s contributing 19, 17, and 12%, respectively. Despite the prominent contribution of individual DBTO_2s, the whole class brings only 9% of the total sulfur. Each class is represented by a large number of individual compounds such that baseline separation of all the compounds present in the mixture is not possible with the current technology. However, we achieved sufficient separation to allow us to assess contributions of classes of compounds, homologous isomers within these classes, and several individual dibenzothiophene sulfones.