

First-principles thermochemistry for gas phase species in an industrial rutile chlorinator

Raphael Shirley¹ Weerapong Phadungsukanan¹ Markus Kraft¹ Jim Downing² Nick Day² Peter Murray-Rust²

released: July 21, 2010

¹ Department of Chemical Engineering and Biotechnology
University of Cambridge
New Museums Site
Pembroke Street
Cambridge, CB2 3RA
UK
E-mail: mk306@cam.ac.uk

² Department of Chemistry
University of Cambridge
Lensfield Road
Cambridge, CB2 1EW
UK

Preprint No. 97



UNIVERSITY OF
CAMBRIDGE

Keywords: titanium tetrachloride, thermochemistry, density functional theory, chlorinator

Edited by

Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
New Museums Site
Cambridge CB2 3RA
United Kingdom

Fax: + 44 (0)1223 334796

E-Mail: c4e@cam.ac.uk

World Wide Web: <http://como.ceb.cam.ac.uk/>



Abstract

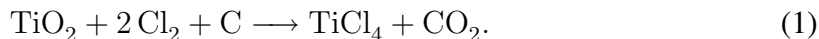
This work presents thermochemical data for possible gas phase intermediate species in an industrial rutile chlorinator. An algorithm developed for previous work is employed to ensure that all possible species are considered, reducing the number of important species neglected. Thermochemical data and enthalpies of formation are calculated for 22 new species using density functional theory, post Hartree-Fock coupled cluster calculations and statistical mechanics. Equilibrium calculations are performed to identify whether any Ti/C intermediates are likely to be important to the high temperature industrial process. These new species are not present at high concentration in the exit stream. It is therefore likely that the two chemical processes do not interact. Rather, the Cl_2 rapidly reacts with the solid TiO_2 forming TiCl_4 and O_2 . The latter then reacts with the solid C forming CO and CO_2 and providing the heat. Data for all the new species is provided as Supplemental Material. Finally, a new methodology for data collaboration is investigated in which the data is made openly accessible using the Resource Description Framework. Example scripts are provided to demonstrate how to query and retrieve the data automatically.

Contents

1	Introduction	3
2	Computational Method	4
2.1	Species Generation	4
2.2	Quantum Chemistry Calculations	6
2.3	Statistical Mechanics and Equilibrium Composition	6
2.4	The internet RDF database	7
3	Results and Discussion	8
3.1	Geometries	8
3.2	Enthalpy of Formation	8
3.3	Thermochemistry	10
3.4	Equilibrium Composition	10
3.5	The internet RDF database	13
4	Conclusions	16
5	Acknowledgments	17
	References	18

1 Introduction

Titanium dioxide (TiO_2) is widely used as a pigment, as a catalyst support, and as a photocatalyst. The combustion synthesis of TiO_2 nanoparticles is a multi-million tonne per year industrial process [5] requiring pure TiCl_4 . This pure TiCl_4 is produced by chlorinating impure TiO_2 [7, 26]. The overall stoichiometry of this ‘chlorination’ process is



Although the chloride process is a mature technology, which has been used in industry since 1958, understanding of the gas-phase reactions of TiCl_4 in a flame remains incomplete [15]. Recently, West *et al.* have developed a detailed kinetic model for the chloride process using an oxygen plasma, including thermochemical data for the titanium oxychloride species involved [29, 30, 39–41]. However, very little is known about the chlorination process. In order to be able to develop a detailed chemical model of the rutile chlorinator thermochemical data must first be obtained.

When investigating new chemical systems there is typically very little information about the chemical species that might be important to the process. It is therefore necessary to screen a large number of molecules. Searching through possible species and conducting tests on them to see if they are relevant is a necessary step towards a full model. This process requires managing large amounts of data. In particular, finding out if species have been investigated in the past involves manually searching old papers and internet databases. Results are often recalculated and important data ignored. For this reason there is a need to improve the current system for storing and distributing data [16].

The main aim of this study is to provide gas-phase thermochemical data for possible titanium and carbon species containing oxygen, and chlorine ($\text{Ti}_i\text{C}_j\text{O}_k\text{Cl}_l$). This is done in order to determine whether the gas phase kinetics will be important in the heterogeneous reaction of Cl_2 with the solid TiO_2 or whether it is possible to neglect the presence of carbon in any subsequent kinetic model. This data has been calculated from first principles using density functional theory (DFT) and statistical mechanics for an exhaustive list of 22 species. In order to ascertain the importance of particular species, equilibrium calculations have been used to identify the most abundant species in an industrial environment. These species may also be important in the flame synthesis of TiO_2 , and further equilibrium simulations are conducted of these. This is motivated by earlier work that raised the possibility of carbon containing species playing an important role [34].

The second aim is to investigate and develop a new method for disseminating thermochemical data. Over the last decade chemical informatics has progressed at a higher rate than it has been adopted by non-specialist chemists and engineers [9, 12, 20–22, 33]. Collating thermochemical data is often an arduous process of manually searching a number of public databases and then converting formats. Here, we aim to distribute the generated data in a manner that is in line with the latest advances in the field. A web server is set up allowing access to the full database of species stored in a resource description framework (RDF) database. The chemical markup language (CML) is updated to include thermochemical data in NASA form allowing free access to the data in computer readable format. Some examples are provided of how this data storage format offers considerable

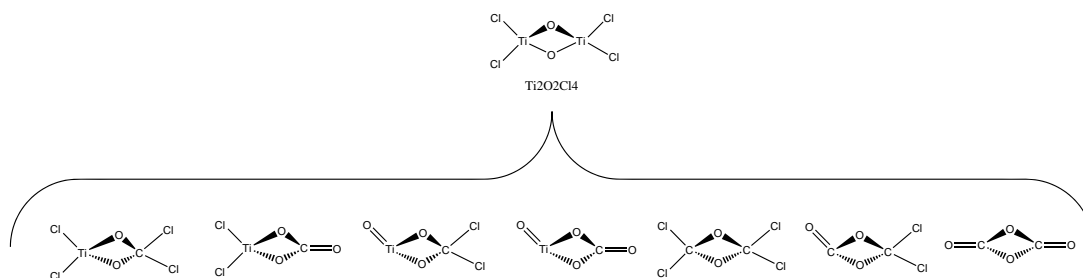


Figure 1: Generating possible hydrogen-containing species from $\text{Ti}_2\text{O}_2\text{Cl}_4$. Thermochemical data for all of these species is provided as Supplemental Material.

advantages over the existing standard whereby data is stored and distributed in single text files in human readable form.

2 Computational Method

2.1 Species Generation

No reaction scheme is available for the chlorination of TiO_2 to form TiCl_4 . It is therefore necessary to determine those species that could possibly be formed in such a system. This section describes the method we have used to generate an exhaustive list of all likely intermediate species. We have recently developed a script to automatically generate species from a selection of existing species [24, 29]. The specific method used here is shown in Algorithm 1. This depends on having access to a simple rule based cleaning function to take the bond lengths to a good first guess. We used the clean function as implemented in the Materials Studio software package [1].

Specifically, the method used here involved replacing titanium atoms with C atoms on the full set of titanium oxychloride ($\text{Ti}_x\text{O}_y\text{Cl}_z$) species from West *et al.* 2009 [41]. It also considered replacements of two chlorine atoms with a double bonded oxygen atom and vice versa. Because the operations involved replacing atoms of the same valence there should be no change of spin after any operation in this algorithm. We therefore used the same spin multiplicity as calculated for the original species of West *et al.* [39, 41]. However, in order to be confident this was a legitimate assumption, single point energy calculations were performed on all of the final geometries to determine if the the lowest energy spin state was different to the initial guess. The assumption held to be valid in every case.

Every possible combination of titanium replacements in the original $\text{Ti}_x\text{O}_y\text{Cl}_z$ species is considered. Figure 1 shows the result of applying this algorithm to the dimer species $\text{Ti}_2\text{O}_2\text{Cl}_4$. In order to remove any duplicates, an International Chemical Identifier (InChI) name, which can uniquely identify molecules of the same connectivity, is assigned to each species using OpenBabel [12, 23]. Duplicate species can then be automatically removed from the set by simply comparing the InChI names. InChI names do not work properly for metallic compounds because they do not allow bonds to metallic atoms. However, they

Algorithm 1: Overview of algorithm used to generate possible Ti/C species

input : connectivity matrices/geometries for original species

output: connectivity matrices/geometries for all new species

```
for input species 1 to input species n do
  import connectivity matrix and geometry for species;
  create temporary list of new species;
  save species to new temporary list;
  count titanium atoms in species and number them;
  for titanium 1 to titanium m do
    for temporary list item 1 to temporary list item l do
      import duplicate of list item;
      replace titanium atom in duplicate list item with a carbon atom;
      clean geometry and save duplicate as new item in temporary list;
      if titanium bonded to two chlorine atoms then
        import duplicate of list item;
        replace two chlorine atoms in duplicate with double bonded oxygen;
        clean geometry and save duplicate as new item in temporary list;
        replace titanium atom in duplicate list item with a carbon atom;
        clean geometry and save duplicate as new item in temporary list;
      end
      if titanium double bonded to oxygen then
        import duplicate of list item;
        replace oxygen in duplicate with two single bonded chlorine atoms;
        clean geometry and save duplicate as new item in temporary list;
        replace titanium atom in duplicate list item with a carbon atom;
        clean geometry and save duplicate as new item in temporary list;
      end
    end
  end
  add temporary list to output list;
  delete temporary list;
end
go through output list and delete species with duplicate connectivity matrices;
```

remain unique and can still be used to check for connectivity duplicates. In principle this might have led to the removal of species that were not duplicates however by checking all removals we confirmed that this was not the case here. This is however, a serious problem with the InChI specification and should be improved.

Each step of this process is fully automated using a combination of Python and Perl scripts. All the new species generated were checked manually both before and after geometry optimization to ensure no important species were absent. Conversely, this check also ensured that physically unrealistic species, such as species where the molecule had clearly broken into two separate molecules under geometry optimization, were identified and removed.

2.2 Quantum Chemistry Calculations

Geometry optimization and molecular properties for all new species were calculated using density functional theory (DFT). Previous studies on the $\text{Ti}_x\text{O}_y\text{Cl}_z$ system [29, 36, 39, 41] indicate that the B97-1 [13] functional is the most reliable and accurate. Consequently, we have used this functional throughout the present study. After geometry optimization with DFT, single point calculations were performed on the calculated geometries using coupled cluster calculations with single, double, and triple excitations, CCSD(T) [27]. These calculations are considered to be more accurate than DFT for this system [39] and give us greater confidence in the calculated enthalpies of formation.

The basis set used for all calculations was 6-311+G(d,p). This consists of the 6-311G basis set [17] for oxygen; the McLean-Chandler (12s,9p) (621111,52111) “negative ion” basis set [19] for chlorine; the Wachters-Hay all-electron basis set [14, 35] for titanium, using the scaling factors of Raghavachari and Trucks [28] and supplementary polarization functions and diffuse functions. Such a supplemented, triple- ζ basis set should be large enough to ensure that the basis set truncation error is comparable with, or smaller than, the inherent errors in the DFT [2]. This is the largest basis set available in GAUSSIAN03 for these elements but may be the main source of error with the CCSD(T) calculations [39].

All calculations were carried out using the GAUSSIAN03 package [8]. The convergence threshold for the self consistent field was set to 10^{-8} Hartrees. The geometry optimization was considered to have converged if the root-mean-squared force on every atom was under 10^{-5} Hartrees/Bohr. Some species underwent a connectivity change after optimization. It was therefore necessary to perform a further duplicate check.

2.3 Statistical Mechanics and Equilibrium Composition

Heat capacities (C_p°), thermal enthalpy ($H(T) - H(0 \text{ K})$), and entropies (S°) were calculated for temperatures in the range 20-4000 K using the rigid rotator harmonic oscillator (RRHO) approximation, taking unscaled vibrational frequencies and rotational constants from the B97-1/6-311+G(d,p) calculations. The contribution of the excited electronic states to the partition function is ignored as in previous studies [29, 39, 41], where tita-

mium containing-species have been shown to not possess low energy excited states.

Polynomials in the NASA [11] form were fitted to $C_p^\circ(T)/R$, H° and S° over the temperature ranges $20-T_x$ K and T_x-4000 K, constrained to ensure that all three functions are continuous and smooth across the boundary temperature, T_x , which was varied to optimize the fit. Using these NASA polynomials, the equilibrium composition as a function of temperature was calculated using the open source software Cantera [10].

2.4 The internet RDF database

In the past thermochemistry was typically calculated by first performing expensive quantum calculations on given initial molecular geometries. Important information was then extracted and stored in a database such as a relational database commonly known as structured query language (SQL) database. Finally, the data was queried and post-processed to calculate important thermochemical quantities.

A relational database stores data in a two-dimensional matrix, known as a table. The data can be accessed through a well-developed structured query language (SQL). Currently, the majority of databases employ such a relational database because of its ease of use and the rapidity of software development in such an environment. However, there is a large amount of information that one can obtain from quantum calculations. This chemical information is multidimensional and cannot be presented in its entirety using a relational database model. Using a relational database with a high dimensional data model only increases the difficulty of our problem, because it involves using an ever larger number of tables with ever increasing redundancy. In order to create a more complete chemical database, a new approach is required.

The fundamental principles of the new database management system can be categorized into three major components.

Data structure and data storage The first major advance in our chemical data storage system is to make use of the Chemical Markup Language (CML) developed by Murray-Rust et al. [22] which is widely used for crystallographic data, polymers, reactions, etc. The CML schema was developed based on the eXtensible Markup Language (XML) standard. In XML, information is represented in a tree-like data structure allowing one to represent chemical data in its native multidimensional format. Also, XML data can be easily extended and comprehended by both human and machine.

Database querying In order to query an XML database an index system and query language must be developed. The graph structure of all the information can be specified using subject, predicate, object statements. These “triples” then give a complete description of the data along with the relations between data points. A method for translating the data in CML format into a list of these triples has been developed which naturally follows the structure of a typical quantum calculation. Queries can then be resolved by walking the data graph.

The standard we use for representing these triples is called the Resource Descriptive Framework (RDF) [4]. Using RDF the graph model can be stored, indexed and

queried. Each node and each relationship in the graph is uniquely identified by a Uniform Resource Identifier (URI) forming a set of RDF triples statements. When the RDF is read by a ‘triplestore’ database (e.g. ‘OpenRDF’, ‘Jena’) it is ready for searching and manipulating using a graph pattern matching query language such as SPARQL or SeRQL.

Data transaction The major benefit of an XML and RDF based data management system is that it provides a platform and protocol independent framework to exchange data. Data can be queried, downloaded and uploaded through a simple ‘REST’ interface over the World Wide Web and one can develop one’s applications in any programming language. This enables open-access to scientific data and data collaboration in the scientific community.

3 Results and Discussion

3.1 Geometries

Converged geometries and molecular properties were found for 22 new species. All this information in the form of GAUSSIAN03 output files, MDL Molfiles, and CML files is provided in the Supplemental Material where species are identified by the unique SMILES name [37, 38] generated using OpenBabel [12]. For a small number of DFT calculations the self-consistent field (SCF) did not converge within 400 iterations. Such species were investigated manually and in all such cases were deemed to be physically unreasonable. Typically they involved two atoms within half an Ångstrom of each other or involved one species which had split into two molecules but retained the now unphysical multiplicity.

In cases where multiple isomers exist for a given empirical formula, all isomers are included. This is because before running equilibrium composition calculations there is no straightforward way to determine which species are more important. Additionally, since they will take part in different reaction paths a future kinetic model may need to consider each isomer individually.

3.2 Enthalpy of Formation

The standard enthalpy of formation at 0 K, $\Delta_f H_{0\text{K}}^\circ$, can be determined by subtracting the energy of the species from the energy of its elements in their standard states as calculated using DFT and then making thermal corrections to the enthalpy by using statistical mechanics. This method often introduces significant errors so the choice of reaction is often critical. In the past, studies into these systems have used both a single standard reaction [29], where very little experimental data is available and isodesmic reactions [24] where more relevant literature data is available. Table 1 shows all the literature values for species that could be used in forming reactions. It is relatively sparse and some of this data is quite old. In addition to that the species are not electronically similar to those species that we are investigating here. For this reason we decided to use a single standard reaction for the

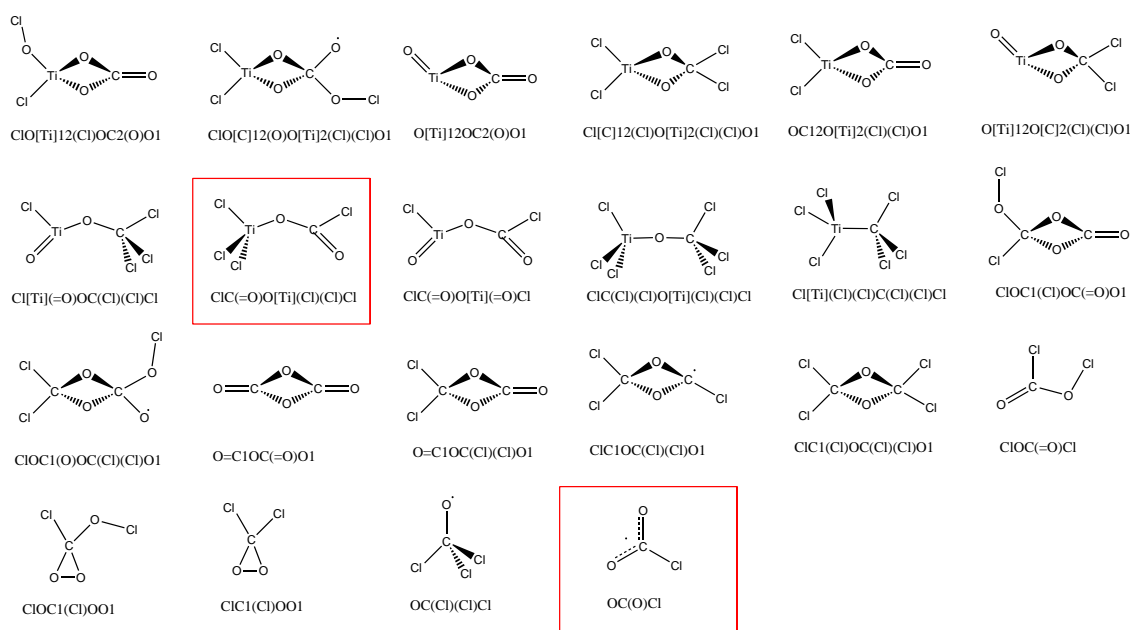
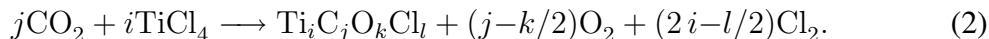


Figure 2: Geometries of all the new species. Precise geometries are available in the Supplemental Material. The Supplemental Material is all organized in terms of the SMILES system for naming molecules, which are reported here so the reader can successfully navigate the Supplemental Material. The species highlighted by red rectangles are those found at highest concentrations in the equilibrium study and correspond to the species in Table 2.

Table 1: Sparse literature values for standard enthalpies of formation at 298.15 K for species that could be used to generate enthalpies of formation.

Species	$\Delta_f H_{298.15\text{ K}}^\circ$ kJ/mol	Reference
TiCl ₄	-763.16	[3]
TiO ₂	-305.43	[3]
TiOCl ₂	-598	[39]
CO ₂	-393.51	[3]
COCl ₂	-220.37	[3]
CCl ₄	-95.60	[3]
CCl ₃	71.10	[3]
CCl ₂	230.10	[3]
CCl	439.57	[3]
Cl ₂	zero	
O ₂	zero	

calculation of Enthalpies of formation,



In order to test the validity of this method we used it to calculate some enthalpies for known species. These results differed from the experimental values by up to 100 kJ/mol. At high temperature, where entropy is the dominant contribution to the Gibbs energy, the equilibrium compositions are relatively insensitive to the enthalpy of formation. While errors of up to 100 kJ/mol in the enthalpy are significant, they are expected and are broadly consistent with those obtained by Phadungsukanan *et al.* [24] for silica species where isodesmic reactions were used. In order to improve these results we conducted further calculations with the expensive but accurate coupled cluster method. These CCSD(T) results did reduce errors in our calculation of known species to a maximum error of 60 kJ/mol. Enthalpies of formation calculated using CCSD(T) differed to those from the DFT results by an average of 20 kJ/mol across all the new species presented. Furthermore, the improved enthalpies of formation had little impact on the equilibrium compositions (mol fraction of highest concentration species at 2000 K changed by < 10 %) and had no effect on the conclusions drawn.

3.3 Thermochemistry

A full list of computed molecular and standard thermochemical data for species convergent under geometry optimization using the B97-1 functional with the 6-311+G(d,p) basis set is presented in the Supplemental Material. Table 2 shows this data for the two most abundant new species in the equilibrium composition (Section 3.4).

3.4 Equilibrium Composition

Equilibrium compositions were calculated using NASA polynomials for the 22 new species using results calculated in this work, those from the titanium oxychloride reaction mech-

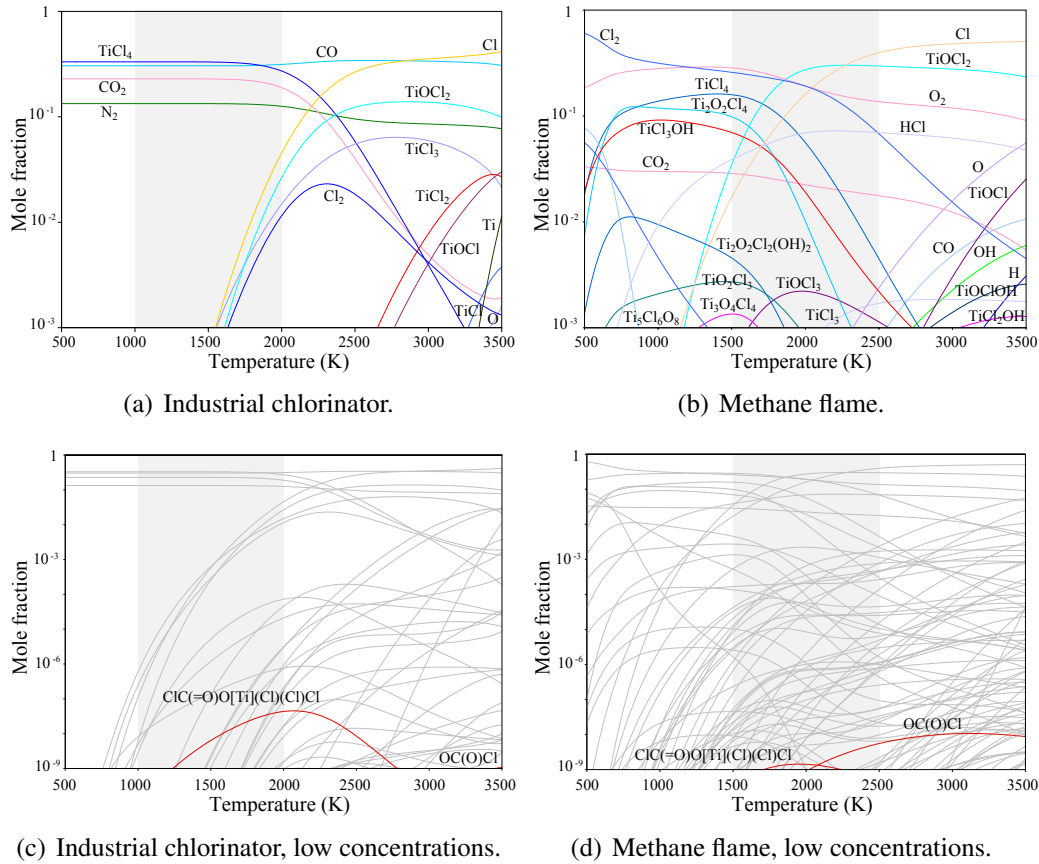


Figure 3: *Computed equilibrium compositions. Figure a) shows equilibrium compositions in a rutile chlorinator based on an initial composition containing 33.2 mol % TiCl_4 , 22.9 mol % CO_2 , 30.5 mol % CO , and 13.4 mol % N_2 at 3 bar. Figure b) shows equilibrium composition in a methane flame based on an initial composition of 48.5 mol % TiCl_4 , 3 mol % CH_4 , and 48.5 mol % O_2 at 1 bar. The industrially relevant conditions are highlighted in both figures by the grey region. Figures c) and d) show compositions for the same simulations but showing lower concentrations and with all species apart from the new species shown in grey.*

Table 2: Standard entropy at 298.15 K, $S_{298.15\text{ K}}^{\circ}$, and molecular properties for the two new species with a max equilibrium concentration above a mole fraction of 10^{-9} calculated at the B97-1/6-311+G(d,p) level of theory for stable electronic ground states. The electronic degeneracy, g , and the rotational symmetry, s , are included. To find corresponding information in the Supplemental Material use the SMILES names provided in Figure 2.

species	$S_{298.15\text{ K}}^{\circ}$ J/mol K	g	s	rot. const. GHz	vibrational frequencies cm^{-1}
OC(O)Cl	280	2	2	13.655 5.1835 3.7573	282.84, 427.62, 644.55, 652.62, 1149.32, 1394.69
ClC(=O)O[Ti](Cl)(Cl)Cl	481	1	1	0.9856 0.4666 0.4380	8.05, 17.24, 30.93, 113.51, 117.92, 121.1, 150.77, 183.86, 292.71, 420.11, 485.13, 499.64, 500.76, 578.41, 672.61, 847.37, 1151.08, 1864.81

anism presented by West *et al.* [41], and all of the species containing only C, O, and N from the GRI-Mech methane mechanism [31]. The West *et al.* reaction mechanism includes NASA polynomials for TiCl_x ($x = 1, \dots, 4$), TiO , TiO_2 , CCl_4 , CCl_3 , CCl_2 , CCl , C , COCl_2 , COCl , chlorine oxides and standard gases, taken from the NASA database [18]. For species which exist in both the GRI-Mech mechanism and the West *et al.* mechanism, the data used is taken from West *et al.*.

Computed equilibrium data calculated over the temperature range of 500–3500 K are shown in Figure 3 for an initial composition containing 33.2 mol % TiCl_4 , 22.9 mol % CO_2 , 30.5 mol % CO , and 13.4 mol % N_2 . These are typical concentrations present in the exit stream of an industrial reactor. The input stream obviously contains no C but is around 95 mol % chlorine and 5 mol % air at 200 °C and 1.8 barg. It is then rapidly heated as it passes through the coke bed. All of the new species presented do not have equilibrium concentrations above a mol fraction of 10^{-4} . In total 2 of the new species have a maximum mol fraction above 10^{-9} over the temperature range 500–3500 K at a pressure of 3 bar. Given the method used to find these new species we can be confident that all stable species will have been included. It is therefore sensible to infer that there is negligible gas phase chemical interaction between carbon and titanium in industrial chlorinators.

Geometries of the new species presented are shown in Figure 2. SMILES format names are given to help the reader associate these geometries with the SMILES used in the Supplemental Material. In an industrial context, where the gases are essentially at equilibrium because surface reaction is rate-limiting, these plots give insight in to the reactions that are likely to be important on the surface. Cl_2 attacking the TiO_2 surface is likely to be the important pathway to formation of gaseous TiCl_4 .

One reason for the relative instability of these Ti/C species might be the high tension associated with a double oxygen bridge. The C–O bond (1.4 Å) is much shorter than the Ti–O bond (1.8 Å). This means that the double oxygen bridges, common in other Ti systems [29, 34], here force the Ti bonding orbitals away from their preferred tetragonal shape and lead to a higher energy. Interestingly, the direct Ti–C bond seems to be ex-

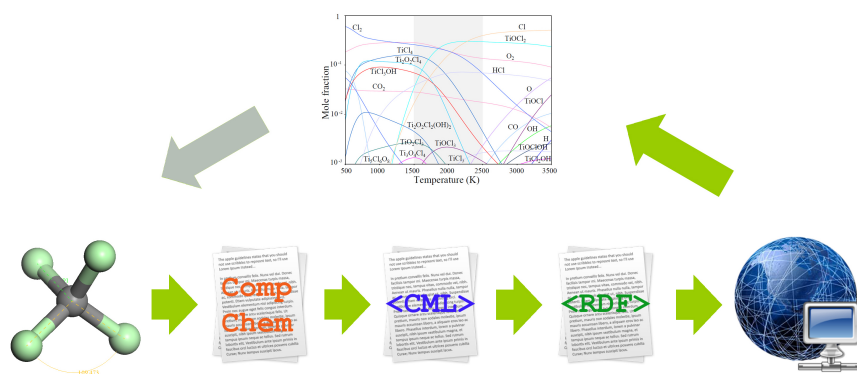


Figure 4: *An overview of the database framework. Geometries for possible species are generated based on chemical insight. Geometries are optimized and properties calculated using quantum chemistry. Data is converted to CML format and in turn RDF. Data is uploaded to internet on database which can be queried. Relevant molecules automatically selected and used to run simulations. Results from simulations can be used to inform further geometries or further calculations to be run.*

tremely unstable, with only TiCCl_6 relaxing to a minimum in the potential energy surface. In general all of the species containing both carbon and titanium will be very short lived. The new CO_2Cl radical is of comparable concentration to some of the established carbon oxychlorides and may be important in other chemical systems.

In addition to industrial rutile chlorinators. There has been some interest in the possibility of carbon containing species playing an important role in industrial flame reactors for producing TiO_2 nanoparticles. Specifically, in recent work by Totton *et al* [34] It was found that hydrogen containing species are stable and will play a role in the TiCl_4 oxidation kinetics. However, it is not clear whether or not carbon containing species play a similar role. For this reason we repeated the simulations in that work including the new species presented here. As in the chlorinator the new species were not found at high concentration and therefore can be safely ignored from any future kinetic model. Figure 3(b) shows the equilibrium results for that system including these new species. The results are indistinguishable from the simulations carried out without these species (Figure 2 [34]). This means the only influence of using a flame instead of an oxygen plasma is the presence of residual H_2O . The presence of CO_2 does not impact kinetics.

3.5 The internet RDF database

In this work, we illustrate the use of an XML/RDF database for storage and distribution of quantum calculation results. We have set up a web server [25] providing open access to this database. This database is managed using OpenRDF which implements the standard SPARQL (querying language) protocol and standard http interface. This allows the use


```

<?xml version="1.0" encoding="UTF-8"?>
<cml xmlns="http://www.xml-cml.org/schema">
  ① <module role="joblist">
    ③ <identifier convention="chemid:EmpiricalFormula" value="C2Cl4O2"/>
    <identifier convention="chemid:InChI" value="InChI=1/C2Cl4O2/c3-1(4)7-2(5,6)8-1"/>
    <identifier convention="chemid:CanonicalSmiles" value="ClC1(Cl)OC(Cl)(Cl)O1"/>
    ② <module role="job" title="job1">
      <module role="init">
        ④ <parameterList>
          ⑤ <parameter dictRef="cmlqm:cmd.geom">
            ⑦ <scalar dataType="xsd:string">Geometry optimization</scalar>
          </parameter>
          <parameter dictRef="cmlqm:qm.basis">
            <scalar dataType="xsd:string">6-311+G(d,p)</scalar>
          </parameter>
        </parameterList>
        <molecule formalCharge="0" spinMultiplicity="1">
          ⑧ <atomArray>
            ⑪ <atom id="a1" elementType="C" x3="-1.663" y3="-1.148" z3="-0.058"/>
            </atomArray>
            <bondArray>
              ⑫ <bond atomRefs2="a1 a2" order="S" id="a1_a2"/>
            </bondArray>
          </molecule>
        </module>
        ⑥ <module role="final">
          ⑨ <molecule formalCharge="0" spinMultiplicity="1">
            <atomArray>
              <atom id="a1" elementType="C" x3="-1.662" y3="-1.151" z3="-0.058"/>
            </atomArray>
            <bondArray>
              <bond atomRefs2="a1 a2" order="S" id="a1_a2"/>
            </bondArray>
            <propertyList>
              ⑬ <property dictRef="cmlqm:property.hf298">
                ⑭ <scalar units="units:kcal.mol-1" ataType="xsd:double">
                  -58.4476055698
                </scalar>
              </property>
            </propertyList>
          </molecule>
        </module>
      </module>
    </module>
  </cml>

```

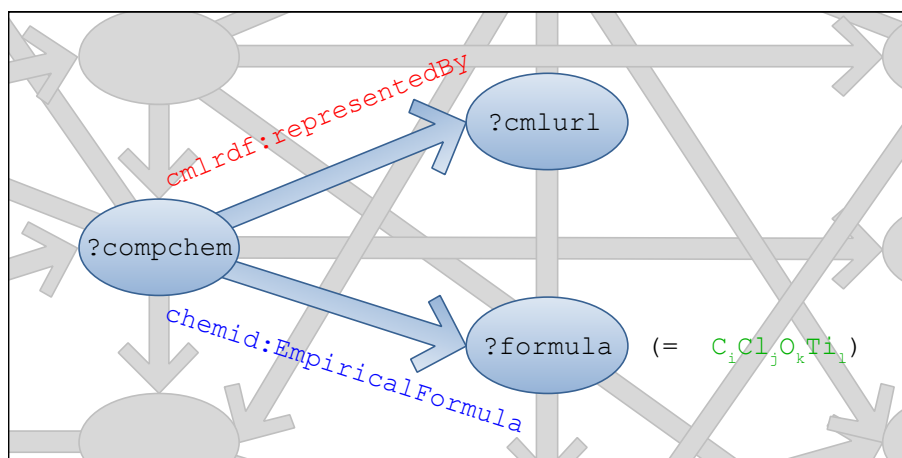
Figure 5: The structure of CML for storing computational chemistry output: (1) CML container, (2) job list module containing multiple computational jobs, (3) job module, (4) initialization of the computation for a job, (5) final output of the computation, (6) parameter list for initialization, (7) initial geometry, charge and multiplicity, (8) optimized geometry, (9) parameter definitions, (10) array of atomic coordinates, (11) array of bonds, (12) property list from the calculations, and (13) property definitions.


```

PREFIX cmlrdf:<http://www.xml-cml.org/rdf-schema#>
PREFIX chemid:<http://www.xml-cml.org/chemid#>
SELECT DISTINCT ?cmlurl
WHERE {
    ?compchem    cmlrdf:representedBy    ?cmlurl;
                chemid:EmpiricalFormula ?formula
FILTER regex(str(?formula),
"(C[0-9]*) (Cl[0-9]*)?(O[0-9]*)?(Ti[0-9]*)?$",
"i")
}

```

(a) SPARQL query.



(b) Graph pattern corresponding to SPARQL query

Figure 6: a) An example SPARQL query which will return URLs for all species which necessarily contain carbon and optionally contain chlorine, and/or oxygen, and/or titanium. It does this by querying the empirical formula and filtering only those species fulfilling this criteria. b) The corresponding graph pattern that would be matched by the SPARQL query.

of common libraries such as ‘SPARQLWrapper’ in Python, ‘Jena’ in Java to connect to the database over the internet. The Supplemental Material contains a number of examples of standard queries. This capability makes the database extremely versatile and could be useful for projects such as RMG [32] and PrIme [6].

The database management system is developed in Java and runs on a Tomcat server. The Java software, which was started in previous work [24], and has been heavily extended for this work is used to parse the Gaussian files, write the results to CML, and then use an RDF ontology developed for computational quantum chemistry to create the RDF statements (triples). This Java package is used to manage the resource files. After the resource files have been put in the filesystem, the RDF is then pushed to the OpenRDF Sesame server which stores and indexes the triples. OpenRDF then executes the SPARQL queries and returns their results.

As an example we have provided a script which queries the database for all species con-

taining carbon necessarily and the elements chlorine, and/or oxygen, and/or titanium optionally, and adds these to an existing thermochemistry file for species manually collected from other databases (as detailed in Section 3.4). The thermochemistry was then identical to that used to generate Figure 3. Indeed, it was simple to modify the script to exactly replicate Figure 3. This script uses the SPARQL query shown in Figure 6. If all thermochemistry databases were available in this manner it would allow mechanism development to proceed at a faster rate. We strongly recommend this kind of online database to other organizations that produce and distribute thermochemical data. The database could be used for tasks such as:

- Select all results produced by a given basis set. Useful for comparing methods.
- Search for all species with a given functional group. Could be used for building a group additivity scheme.
- Suggest possible reactions by matching functional groups. Could be used to automatically generate reaction mechanisms.

4 Conclusions

This work extends the detailed thermochemistry of the gas-phase titanium oxychloride species investigated by West *et al.* [39] to a similar system of carbon-containing species. New molecular geometries have been proposed through automated species generation and investigated using density functional theory with the B97-1 functional and the 6-311+G(d,p) basis set. Further calculations using CCSD(T) were then used to improve values for the enthalpy of formation. This enabled detailed thermochemical data to be computed for the generated species, which have not been reported in the literature and are difficult to obtain experimentally.

Using the thermochemical data, temperature-dependent equilibrium compositions were obtained through simulations at industrial operating conditions. These clearly show that none of the new species containing both titanium and carbon are found at high concentrations. The work therefore demonstrates the possibility of an automatic screening process. More general algorithms could be developed in the future allowing a list of all possible chemical species to be generated automatically. This would then allow important species to be investigated in more detail. Finally, the new thermochemical data is made available in an online RDF database allowing machine querying and access. To summarize:

- There is very little chemical interaction between Ti and C in industrial rutile chlorinators. The gaseous chlorine is likely to directly react with the solid TiO_2 surface forming TiCl_4 , possibly via intermediate $\text{Ti}_x\text{O}_y\text{Cl}_z$ species.
- Furthermore, these $\text{Ti}_i\text{C}_j\text{O}_k\text{Cl}_l$ species do not play an important role in the flame synthesis of TiO_2 . The presence of residual water in the flame is the only important impact of using a flame to heat the reactants.

- Using an open RDF database for the dissemination of thermochemical data allows more efficient algorithms for investigating new chemical systems. Making the wealth of currently existing data available in this way will make development of kinetic models significantly faster and more reliable.

5 Acknowledgments

The authors thank the EPSRC for help in the form of research hopping grant number EP/E01724X-1 and TiOxide Europe Limited for the financial support of RS, and The Cambridge Commonwealth Trust for the financial support of WP.

References

- [1] Accelrys. Materials studio v4.3, 2008. URL <http://accelrys.com/products/materials-studio/>.
- [2] A. D. Boese, J. M. L. Martin, and N. C. Handy. The role of the basis set: Assessing density functional theory. *J. Chem. Phys.*, 119(6):3005–3014, Aug 2003. ISSN 0021-9606. doi:10.1063/1.1589004.
- [3] M. W. Chase Jr. NIST-JANAF thermochemical tables, fourth edition. *J. Phys. Chem. Ref. Data*, Monograph 9:I–II, 1998.
- [4] W. W. W. Consortium. Resource description framework (rdf), 2004. URL <http://www.w3.org/RDF/>.
- [5] J. Emsley. *Molecules at an Exhibition*. Oxford University Press, 1999.
- [6] M. Frenklach. Transforming data into knowledge—process informatics for combustion chemistry. *Proceedings of the Combustion Institute*, 31(1):125 – 140, 2007. ISSN 1540-7489. doi:10.1016/j.proci.2006.08.121.
- [7] W. Frey. Process for the production of titanium tetrachloride. US Patent 2,790,703, Apr 1957.
- [8] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02, 2003. Gaussian, Inc., Wallingford, CT, 2004.
- [9] G. V. Gkoutos, P. Murray-Rust, H. S. Rzepa, and M. Wright. Chemical markup, xml, and the world-wide web. 3. toward a signed semantic chemical web of trust. *Journal of Chemical Information and Computer Sciences*, 41:1124 – 1130, 2001. doi:10.1021/ci000406v.
- [10] D. G. Goodwin. An open source, extensible software suite for CVD process simulation. In M. Allendorff, F. Maury, and F. Teyssandier, editors, *Chemical Vapor Deposition XVI and EUROCVI 14*, volume 2003-08 of *ECS Proceedings*, pages 155–162, Pennington, New Jersey, USA, 2003. The Electrochemical Society, The Electrochemical Society. URL www.cantera.org.

- [11] S. Gordon and B. J. McBride. Computer program for calculation of complex chemical equilibrium composition, rocket performance, incident and reflected shocks and chapman-jouguet detonations. Technical Report NASA-SP-273, NASA, Glenn Research Center, Cleveland, OH, 1976. URL <http://hdl.handle.net/2060/19780009781>.
- [12] R. Guha, M. T. Howard, G. R. Hutchison, P. Murray-Rust, H. Rzepa, C. Steinbeck, J. Wegner, and E. L. Willighagen. The blue obelisk interoperability in chemical informatics. *Journal of Chemical Information and Modeling*, 46(3):991–998, 2006. doi:10.1021/ci050400b.
- [13] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy. Development and assessment of new exchange-correlation functionals. *J. Chem. Phys.*, 109(15):6264–6271, October 1998. doi:10.1063/1.477267.
- [14] P. J. Hay. Gaussian basis sets for molecular calculations. the representation of 3D orbitals in transition-metal atoms. *J. Chem. Phys.*, 66:4377–4384, 1977. doi:10.1063/1.433731.
- [15] B. Karlemo, P. Koukkari, and J. Paloniemi. Formation of gaseous intermediates in titanium(IV) chloride plasma oxidation. *Plasma Chem. Plasma Process.*, 16:59–77, 1996. doi:10.1007/BF01465217.
- [16] M. Kraft and S. Mosbach. The future of computational modelling in reaction engineering. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1924):3633–3644, 2010. doi:10.1098/rsta.2010.0124.
- [17] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople. Self-consistent molecular-orbital methods. 20. Basis set for correlated wave-functions. *J. Chem. Phys.*, 72:650–654, 1980.
- [18] B. J. McBride, M. J. Zehe, and S. Gordon. NASA Glenn coefficients for calculating thermodynamic properties of individual species. Technical Report TP-2002-211556, NASA, Glenn Research Center, Cleveland, OH, 2002. URL <http://hdl.handle.net/2060/20020085330>.
- [19] A. D. McLean and G. S. Chandler. Contracted gaussian-basis sets for molecular calculations. 1. 2nd row atoms, z=11-18. *J. Chem. Phys.*, 72:5639–5648, 1980.
- [20] P. Murray-Rust and H. S. Rzepa. Chemical markup, xml, and the worldwide web. 1. basic principles. *Journal of Chemical Information and Computer Sciences*, 39:928–942, 1999. doi:10.1021/ci990052b.
- [21] P. Murray-Rust and H. S. Rzepa. Chemical markup, xml and the world-wide web. 2. information objects and the cml dom. *Journal of Chemical Information and Computer Sciences*, 41:1113–1123, 2001. doi:10.1021/ci000404a.
- [22] P. Murray-Rust and H. S. Rzepa. Chemical markup, xml, and the world wide web. 4. cml schema. *Journal of Chemical Information and Computer Sciences*, 43:757–772, 2003. doi:10.1021/ci0256541.

- [23] N. O’Boyle, C. Morley, and G. Hutchison. Pybel: a python wrapper for the openbabel cheminformatics toolkit. *Chemistry Central Journal*, 2(1):5, 2008. ISSN 1752-153X. doi:10.1186/1752-153X-2-5.
- [24] W. Phadungsukanan, S. Shekar, R. Shirley, M. Sander, R. H. West, and M. Kraft. First-principles thermochemistry for silicon species in the decomposition of tetraethoxysilane. *J. Phys. Chem. A*, 113(31):9041–9049, Jul 2009. doi:10.1021/jp905494s.
- [25] W. Phadungsukanan, R. Shirley, and M. Kraft. The Computational Modelling group’s open thermochemistry database. Web database, 2010. URL <http://como.cheng.cam.ac.uk/thermodatabase/>.
- [26] L. Piccolo, B. Calcagno, and M. Ghirga. Method for preparing titanium tetrachloride. US Patent 3,787,556, Jan 1974.
- [27] J. A. Pople, M. Head-Gordon, and K. Raghavachari. Quadratic configuration interaction. a general technique for determining electron correlation energies. *The Journal of Chemical Physics*, 87(10):5968–5975, 1987. doi:10.1063/1.453520.
- [28] K. Raghavachari and G. W. Trucks. Highly correlated systems - excitation-energies of 1st row transition-metals Sc-Cu. *J. Chem. Phys.*, 91:1062–1065, 1989.
- [29] R. Shirley, Y. Liu, T. S. Totton, R. H. West, and M. Kraft. First-principles thermochemistry for the combustion of a TiCl₄ and AlCl₃ mixture. *J. Phys. Chem. A*, 113(49):13790–13796, 2009. doi:10.1021/jp905244w.
- [30] R. Shirley, M. Kraft, and O. R. Inderwildi. Electronic and optical properties of aluminium-doped anatase and rutile TiO₂ from ab initio calculations. *Phys. Rev. B*, 81(7):075111, Feb 2010. doi:10.1103/PhysRevB.81.075111.
- [31] G. P. Smith, D. M. Golden, M. Frenklach, N. W. Moriarty, B. Eiteneer, M. Goldenberg, C. T. Bowman, R. K. Hanson, S. Song, W. C. Gardiner, V. V. Lissianski, and Z. Qin. GRI-Mech 3.0. software package, 2008. URL <http://me.berkeley.edu/gri-mech/>.
- [32] J. Song. *Building Robust Chemical Reaction Mechanisms: Next Generation of Automatic Model Construction Software*. PhD thesis, Massachusetts Institute of Technology, Cambridge MA, USA, February 2004. URL <http://hdl.handle.net/1721.1/30058>.
- [33] K. R. Taylor, R. J. Gledhill, J. W. Essex, J. G. Frey, S. W. Harris, and D. C. De Roure. Bringing chemical data onto the semantic web. *Journal of Chemical Information and Modeling*, 46:939 – 952, 2006. doi:10.1021/ci050378m.
- [34] T. S. Totton, R. Shirley, and M. Kraft. First-principles thermochemistry for the combustion of TiCl₄ in a methane flame. *Proc. Combust. Inst. (In Press)*, 2010. doi:10.1016/j.proci.2010.05.011.
- [35] A. J. Wachters. Gaussian basis set for molecular wavefunctions containing third-row atoms. *J. Chem. Phys.*, 52:1033–1036, 1970. doi:10.1063/1.1673095.

- [36] T.-H. Wang, A. M. Navarrete-Lopez, S. Li, D. A. Dixon, and J. L. Gole. Hydrolysis of TiCl_4 : Initial steps in the production of TiO_2 . *The Journal of Physical Chemistry A*, 114(28):7561–7570, 2010. doi:10.1021/jp102020h.
- [37] D. Weininger. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. and Comput. Sci.*, 28(1): 31–36, 1988. doi:10.1021/ci00057a005.
- [38] D. Weininger, A. Weininger, and J. Weininger. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.*, 29(2):97–101, 1989. doi:10.1021/ci00062a008.
- [39] R. H. West, G. J. O. Beran, W. H. Green, and M. Kraft. First-principles thermochemistry for the production of TiO_2 from TiCl_4 . *J. Phys. Chem. A*, 111(18):3560–3565, 2007. doi:10.1021/jp0661950.
- [40] R. H. West, M. S. Celnik, O. R. Inderwildi, M. Kraft, G. J. O. Beran, and W. H. Green. Toward a comprehensive model of the synthesis of TiO_2 particles from TiCl_4 . *Ind. Eng. Chem. Res.*, 46(19):6147–6156, 2007. ISSN 0888-5885. doi:10.1021/ie0706414.
- [41] R. H. West, R. A. Shirley, M. Kraft, C. F. Goldsmith, and W. H. Green. A detailed kinetic model for combustion synthesis of titania from TiCl_4 . *Combust. Flame*, 156: 1764–1770, 2009. doi:10.1016/j.combustflame.2009.04.011.