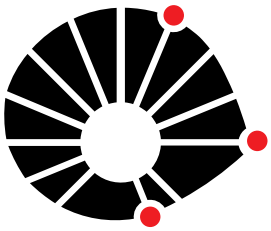

UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Faculdade de Engenharia Química

EMERSON PARAZZI LYRA

Parameters determination for a molecular-based, equation of state from quantum mechanical calculations: towards a predictive equation of state

Determinação de parâmetros para equações de estado baseadas em moléculas a partir de cálculos de mecânica quântica: rumo à uma equação de estado preditiva

Campinas

2023

Emerson Parazzi Lyra

Parameters determination for a molecular-based, equation of state from quantum mechanical calculations: towards a predictive equation of state

Determinação de parâmetros para equações de estado baseadas em moléculas a partir de cálculos de mecânica quântica: rumo à uma equação de estado preditiva

Thesis presented to the School of Chemical Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Chemical Engineering.

Tese apresentada a Faculdade de Engenharia Química da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Química.

Supervisor: Luís Fernando Mercier Franco

Este exemplar corresponde à versão final da Tese defendida pelo aluno Emerson Parazzi Lyra e orientada pelo Prof. Dr. Luís Fernando Mercier Franco.

Campinas

2023

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

Lyra, Emerson Parazzi, 1990-
L995P Parameters determination for a molecular-based, equation of state from quantum mechanical calculations : towards a predictive equation of state / Emerson Parazzi Lyra. – Campinas, SP : [s.n.], 2023.

Orientador: Luís Fernando Mercier Franco.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Química.

1. Cálculos ab initio. 2. Dinâmica molecular. 3. Equação de estado. I. Franco, Luís Fernando Mercier, 1988-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Química. III. Título.

Informações Complementares

Título em outro idioma: Determinação de parâmetros para equações de estado baseadas em moléculas a partir de cálculos de mecânica quântica : rumo à uma equação de estado preditiva
Palavras-chave em inglês:
Ab initio calculations
Molecular dynamics
Equation of state
Área de concentração: Engenharia Química
Titulação: Doutor em Engenharia Química
Banca examinadora:
Luís Fernando Mercier Franco [Orientador]
André Rodrigues Muniz
Marcelo Castier
Caetano Rodrigues Miranda
Frederico Wanderley Tavares
Data de defesa: 12-12-2023
Programa de Pós-Graduação: Engenharia Química

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-7969-3764
- Currículo Lattes do autor: http://lattes.cnpq.br/2985580274093497

Folha de Aprovação da Defesa de Tese de Doutorado defendida por **EMERSON PARAZZI LYRA** e aprovada em 12/12/2023 pela Comissão Examinadora constituída pelos doutores:

Prof. Dr. Luís Fernando Mercier Franco

Presidente e Orientador

FEQ / UNICAMP

(Videoconferência)

Dr. André Rodrigues Muniz

Universidade Federal do Rio Grande do Sul / Porto Alegre – RS

(Videoconferência)

Dr. Marcelo Castier

Texas A&M University at Qatar / Doha

(Videoconferência)

Dr. Caetano Rodrigues Miranda

Instituto de Física, Universidade de São Paulo / São Paulo – SP

(Videoconferência)

Dr. Frederico Wanderley Tavares

UFRJ - Escola de Química / Rio de Janeiro - RJ

(Videoconferência)

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Life is about constancy, insistence, and resistance.
Dedicated to those who never gave up on their dreams.

Acknowledgements

In portuguese.

Primeiramente, eu agradeço a Deus pelo dom da vida. Durante a execução desta tese de doutorado as provações foram muitas (pandemia, troca de orientação, problemas pessoais, desafios do trabalho, ...). Sobrevivi.

Em segundo lugar, eu gostaria de agradecer ao professor Luís Fernando Mercier Franco da Faculdade de Engenharia Química da UNICAMP por toda a orientação dada até a conclusão desta tese de doutorado. Trabalhar no doutorado com o Luís foi uma experiência muito enriquecedora e inspiradora. O Luís gosta do que faz, tem talento e vontade para trabalhar, isso foi fundamental para eu criar forças para superar muitas das situações de dificuldade que apareceram durante o processo. Aprendi e evolui muito. Trabalhei com a pessoa certa. Expresso aqui a minha gratidão ao Luís pelas portas abertas e pela parceria que fizemos até então.

Em terceiro lugar, faço agradecimento a todos os membros convidados para as bancas ao longo da trajetória desta tese de doutorado e que se proporam a dedicar um pouco de seu valioso tempo para contribuir com a evolução deste trabalho. Expresso aqui o meu reconhecimento.

Em quarto lugar, eu gostaria de agradecer algumas pessoas que direta ou indiretamente também foram importantes na trajetória desta tese de doutorado. Faço agradecimento a minha mãe que sempre me incentivou a conclusão dos meus estudos. Faço agradecimento aos familiares mais próximos (tios e primos) que tornaram alguns episódios dessa trajetória mais leves. Faço agradecimento aos camaradas Danilo e Ícaro pela inspiração e pela amizade pura e verdadeira, cultivada desde os primórdios na graduação na Faculdade de Engenharia Química da UNICAMP. Também faço agradecimento ao colega Marco da Faculdade de Engenharia Química da UNICAMP pelo companheirismo, principalmente durante alguns momentos conturbados na pós-graduação. Faço agradecimento aos colegas do Laboratório de Engenharia de Sistemas Complexos da Faculdade de Engenharia Química da UNICAMP, em especial ao Rodrigo e a Marcelle que me auxiliaram muito no início dos trabalhos. Faço agradecimento ao professor Rogerio Custódio do Instituto de Química da UNICAMP, o qual me inspirou a aprender um pouco de Química Quântica. Expresso aqui a minha consideração por vocês.

Não posso deixar de agradecer também a faculdade de Engenharia Química da UNICAMP como instituição, por toda estrutura dada ao longo da minha trajetória acadêmica desde a graduação, percorrendo o mestrado e, agora, rumo a conclusão do doutorado. Já se passaram 13 anos, desde então. Eu agradeço aos professores que me inspiraram profissionalmente (não citarei nomes, pois são muitos...) e demais funcionários da instituição de ensino (principalmente ao pessoal da secretária de pós-graduação) sendo responsáveis por fazerem a "engrenagem" da faculdade funcionar. Expresso aqui a minha gratidão pelos seus serviços prestados.

O presente trabalho foi realizado com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil. Aqui faço meu agradecimento ao CNPq (processo n° 140563/2019−9) pelo financiamento da bolsa de estudos que foi imprescindível para o desenvolvimento das atividades de pesquisa e conclusão desta tese de doutorado.

Este trabalho também foi realizado com apoio da FAPESP, Fundação de Amparo à Pesquisa do Estado de São Paulo - Brasil. Aqui também faço meu agradecimento a FAPESP (processo n° 2018/02713−8) pelo financiamento dos recursos computacionais que auxiliaram o desenvolvimento desta tese de doutorado.

Este trabalho utilizou recursos do "Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-SP)". Aqui faço agradecimento ao CENAPAD-SP pelo acesso aos recursos computacionais que primordiais para o desenvolvimento das atividades de pesquisa e conclusão desta tese de doutorado. Aqui também faço menção de agradecimento ao pessoal do suporte do CENAPAD-SP que sempre foram muito solícitos e rápidos na resolução de dúvidas e problemas que enfrentei no processo de desenvolvimento desta tese de doutorado.

Esta pesquisa utilizou os recursos computacionais e o auxílio do Centro de Computação John David Rogers (CCJDR) do Instituto de Física "Gleb Wataghin" da Universidade de Campinas. Aqui faço agradecimento ao pessoal do CCJDR do Instituto de Física Gleb Wataghin da UNICAMP pelo suporte dado na utilização dos recursos computacionais do Laboratório de Engenharia de Sistemas Complexos da Faculdade de Engenharia Química da UNICAMP.

"What we do in life, echoes in eternity."
General Maximus (in the movie Gladiator)

Abstract

The computation of experimental observables of fluid systems and soft matter using theoretical and computational tools has always drawn attention of the physical, chemical, and engineering sciences. Molecular-based, equations of state are models that have proven to be a powerful tool for this task. The great advantage of molecular-based, equations of state is the direct relation between their parameters and the molecular interactions. One of the most popular families of this type of model is the variants of equations of state derived from the statistical associating fluid theory (SAFT). In SAFT equations of state, molecules are modeled as spherical segments subject to an interatomic potential. Parameters representing interatomic potentials can be obtained from top-down and bottom-up methodologies. In the top-down methodology, the SAFT equations of state parameters are fitted directly to macroscopic experimental data. In the bottom-up methodology, interatomic potential parameters are obtained via *ab initio* calculations. Although the top-down methodology route is the most direct to obtain accurate SAFT equations of state parameters, the need to obtain experimental observables to calibrate the model can be costly and time-consuming. Due to the lack of studies in the literature, this work aims at developing a bottom-up methodology to obtain realistic interatomic potential parameters from *ab initio* calculations to give molecular-based, equations of state a purely predictive character. Interaction energy curves were built using *ab initio* calculations by supermolecular and perturbational methods for dimers of different configurations for a set of 26 molecules of different chemical classes. The molecules were modeled according to a spherical 1-site coarse-grained model, and the Mie potential parameters were fitted to the *ab initio* data for use in the SAFT-VR Mie equation of state. Comparing the Mie potential parameters obtained from the bottom-up methodology of this study with those from the top-down methodology available in the literature within the SAFT framework, an important connection between the macro and quantum scales was observed. Parameters of the bottom-up methodology obtained through an average of the Mie potential parameters obtained for each considered dimer configuration presented values very close to those of the top-down methodology. Furthermore, for most of the molecules studied, when the adopted model was considered appropriate, good results were observed in predicting thermophysical properties, critical point, and saturation properties with SAFT-VR Mie equation of state for both top-down and bottom-up-based Mie potential parameters. Although for some molecules the prediction errors of thermophysical properties, critical point, and saturation properties for bottom-up-based Mie potential parameters were unsatisfactory for the use of purely predictive SAFT-VR Mie in the design of chemical plants, the methodology developed in this study is consistent and can be improved with the development of more complex models to represent the studied molecules.

Keywords: *Ab initio* calculations, Molecular dynamics, Equation of state.

Resumo

O cálculo de experimentais observáveis de sistemas fluidos e matéria mole utilizando ferramentas teóricas e computacionais sempre atraiu a atenção das ciências físicas, químicas e de engenharia. As equações de estado baseadas em moléculas são modelos que provaram ser uma ferramenta poderosa para esta tarefa. A grande vantagem das equações de estado baseadas em moléculas é a relação direta entre seus parâmetros e as interações moleculares. Uma das famílias mais populares deste tipo de modelo são as variantes de equações de estado derivadas da *statistical associating fluid theory* (SAFT). Nas equações de estado SAFT, as moléculas são modeladas como segmentos esféricos sujeitos a um potencial interatômico. Parâmetros que representam potenciais interatômicos podem ser obtidos a partir das metodologias *top-down* e *bottom-up*. Na metodologia *top-down*, os parâmetros das equações de estado da SAFT são ajustados diretamente para reproduzir dados experimentais macroscópicos. Na metodologia *bottom-up*, os parâmetros do potencial interatômico são obtidos por meio de cálculos *ab initio*. Embora a rota da metodologia *top-down* seja a mais direta para obter parâmetros precisos das equações da SAFT, a necessidade de se obter experimentais observáveis para calibrar o modelo pode ser custosa e demorada. Devido à falta de estudos na literatura, este trabalho visa desenvolver uma metodologia *bottom-up* para obter parâmetros do potencial interatômico a partir de cálculos *ab initio*, conferindo às equações de estado baseadas em moléculas um caráter puramente preditivo. Curvas de energia de interação foram construídas utilizando cálculos *ab initio* por métodos supermolecular e perturbacional para dímeros de diferentes configurações para um conjunto de 26 moléculas de classes químicas distintas. As moléculas foram modeladas através de um modelo esférico *coarse-grained* de um site e os parâmetros do potencial de Mie foram ajustados aos dados *ab initio* para uso na equação de estado SAFT-VR Mie. Comparando os parâmetros do potencial de Mie obtidos a partir da metodologia *bottom-up* deste estudo com aqueles da metodologia *top-down* disponíveis na literatura no âmbito do SAFT, observou-se uma importante conexão entre as escalas macro e quântica. Os parâmetros da metodologia *bottom-up* obtidos através de uma média dos parâmetros do potencial de Mie obtidos para cada configuração de dímero considerada apresentaram valores muito próximos aos da metodologia *top-down*. Além disso, para a maioria das moléculas estudadas, quando o modelo adotado foi considerado apropriado, bons resultados foram observados na predição de propriedades termofísicas, ponto crítico e propriedades de saturação com a equação de estado SAFT-VR Mie para parâmetros de potencial Mie baseados em ambas as metodologias *top-down* e *bottom-up*. Embora para algumas moléculas os erros de predição de propriedades termofísicas, ponto crítico e propriedades de saturação para parâmetros de potencial de Mie baseados na metodologia *bottom-up* sejam insatisfatórios para o uso da SAFT-VR Mie puramente preditiva no projeto de plantas químicas, a metodologia desenvolvida neste estudo se mostrou consistente e ainda pode ser aprimorada com o desenvolvimento de modelos mais complexos para representar as moléculas estudadas.

Palavras-chave: Cálculos *ab initio*, Dinâmica molecular, Equação de estado.

List of Figures

List of Tables

Contents

1 Introduction

"Bring forward what is true. Write it so that it is clear. Defend it to your last breath."
Ludwig Boltzmann

1.1 Motivation

Chemical industrial process design and operation rely on accurate measurement, prediction, or estimation of thermophysical properties the involved chemical compounds. Experimental determination of the required data can be costly and time-consuming. One of the main objectives of scientists and chemical engineers is to minimize, or even eliminate, the dependence on data obtained experimentally through correlations and predictive models. The development of accurate correlations and predictive models capable of predicting properties of pure compounds and mixtures in a wide range of thermodynamic conditions is of great importance for developing and optimizing the design of chemical processes.

Traditional approaches, such as empirical or cubic equations of state and activity coefficient models, can adequately describe some systems. There are, however, certain limitations to traditional approaches in predicting properties of complex systems. These traditional approaches often fail outside their limited range of application. More advanced techniques can be used to address this challenge [1].

Molecular-based, equations of state are physically based models developed through perturbation theory techniques. The advantage of molecular-based, equations of state is that they consider the underlying molecular nature of the system. Specifically, molecular-based, equations have a reliable predictive and extrapolative power, as they consider different microscopic contributions that influence macroscopic properties. Molecular-based, equations of state incorporate specific physical interactions, resulting in a model with physically meaningful parameters. As a result, molecular-based, equations can offer a more detailed and comprehensive approach to predicting properties of complex systems [1]. One of the most popular models within this approach is the SAFT equation of state, originally proposed by Chapman *et al.* [2, 3] and derived from the statistical associating fluid theory.

The parameterization of equations of state involve balancing the need for accuracy in computed properties with the desire to minimize experimental effort. The equation of state parameters are usually fitted to reproduce experimental property data. As molecular-based, equations of state are built through a more rigorous theoretical foundation that considers the molecular nature of the compounds, the parameters of the equation of state can be obtained

through other means than fitting experimental data. An option that has become viable over the years due to the development of computational capacity is obtaining parameters for molecular-based, equations of state through *ab initio* calculations.

Within the SAFT framework, the literature has limited references that discuss obtaining parameters through *ab initio* calculations [4–10]. Although the topic has been discussed for some time, there is still no consensus in the literature regarding the development of a consistent methodology for obtaining parameters from *ab initio* calculations to be used in molecular-based, equations of state.

The ultimate goal in the last few decades has been to accurately predict macroscopic properties from *ab initio* calculations avoiding any experimental data other than the atomic constitution of the compounds. In the literature, an exploratory window is still open to develop purely predictive molecular-based, equations of state with parameters obtained through *ab initio* calculations. It is within this context that this dissertation serves as a contribution to the field.

1.2 Literature review

In the physical, chemical, and engineering sciences, developing theoretical and computational tools to predict macroscopic properties of fluid and soft matter systems is a relevant and active topic. Literature reviews [11–17] cover the impact and prospects of matter computer simulation methods and discussions on this topic have been recently extended [18]. Predicting macroscopic properties through theoretical-computational tools depends on the time and length scales involved in the studied problem. Fig. 1 illustrates four time and length scales addressed in this study (*i.e.*, quantum, molecular, meso, and macro) and some approaches to transferring information across these scales.

At the quantum scale, matter is described in terms of electronic and nuclear degrees of freedom. This description falls under the domain of quantum mechanics. In quantum mechanics, physical systems are described by the Schrödinger's equation [19]. Most applications aim at solving the non-relativistic, time-independent, and fixed-nuclei Schrödinger's equation to model molecular systems with many electrons and nuclei without external fields.

The dynamics of systems with many electrons and nuclei require sophisticated computational methods to numerically solve the Schrödinger's equation. These methods are commonly referred to *ab initio* (Latin for "from first principles" or "from the beginning"). Hartree-Fock [20–24] and density functional theory [25, 26] constitute the two most important *ab initio* methods. Hartree-Fock method uses the electronic wave function to solve the problem of many electrons and nuclei. In contrast, density functional theory uses electron density as an alternative to the electronic wave function with gains in problem-solving speed.

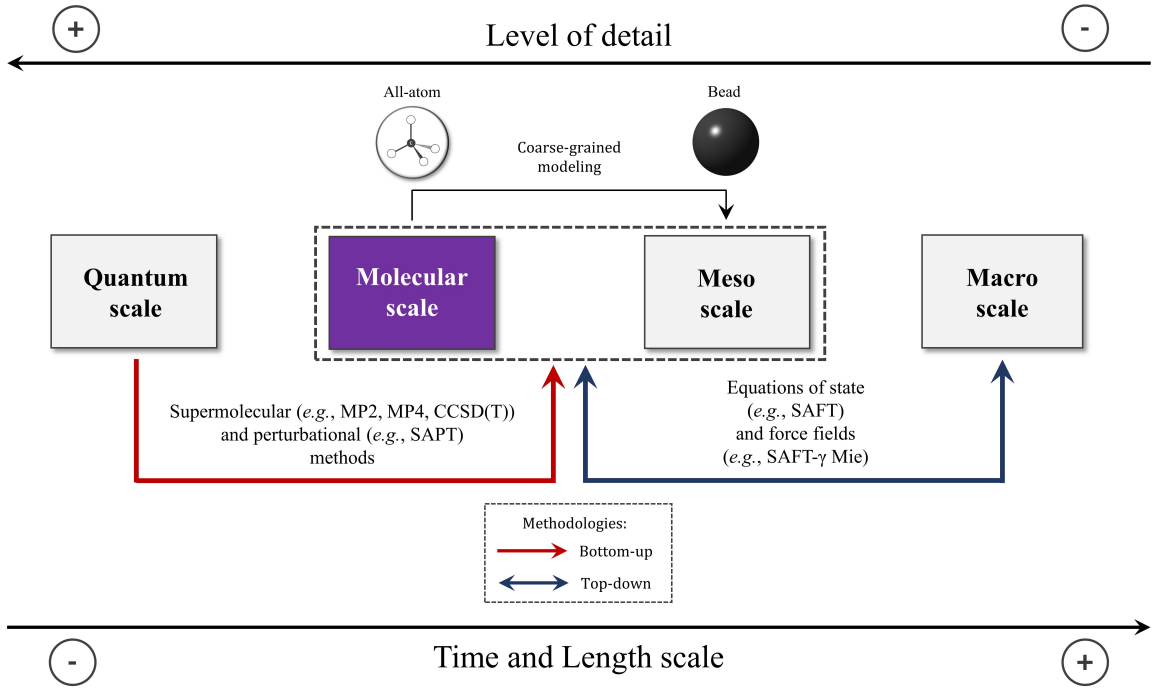

Figure 1 – Examples of length and time scales covered in this work and how information can be transferred across these scales.

While *ab initio* methods can offer comprehensive information for systems of many electrons and nuclei, their computational requirements are substantial. As a result, their use is restricted to systems comprising only a few atoms and short time scales (see Fig. 1). If details of the electronic structure can be suppressed or neglected, the approach to describing matter can be simplified and performed at the molecular level.

At the molecular scale, the dynamics of many-body systems are used for describing matter. For this task, molecular dynamics simulations entail one of the most commonly discussed theoretical-computational methods in the literature [12–15, 27]. In atomistic molecular dynamics simulations, atoms are treated as particles. This simplification makes atomistic molecular dynamics simulations suitable for studying complex molecular systems comprising thousands or millions of atoms in time intervals higher than the quantum scale (see Fig. 1). In molecular dynamics simulations, particles interact through a force field. Force fields are a set of empirical equations that describe the behavior of bonded and non-bonded atoms.

The process of obtaining force field parameters is called force field parameterization, and it can be performed via top-down and bottom-up methodologies. The top-down methodology involves deriving interatomic potential parameters from experimental data. On the other hand, the bottom-up methodology determines interatomic potential parameters from the results of *ab initio* calculations. Top-down-based interatomic potential parameters have greater representability while bottom-up-based interatomic potential parameters generally have greater transferability [12–15, 27].

Determining reliable interatomic potential parameters is one of the biggest challenges in developing force fields for classical molecular simulations. To obtain the parameters of the non-bonded potential, it is essential to understand the nature of intermolecular interactions [12–15, 27]. Supermolecular and perturbational methods are two procedures discussed in the literature to compute the interaction energy in quantum chemistry [28–30]. They are examples of how information can be transferred from the quantum to the molecular scale by developing interatomic potentials (see Fig. 1).

In the supermolecular methods, the interaction energy is computed by subtracting the sum of the energies of individual monomers from the energy of the dimer. The Hartree-Fock method and density functional theory do not accurately describe non-bonded interactions [28–30]. To address this problem, other *ab initio* methods, such as dispersion-corrected density functional theory [31–33], Møller-Plesset perturbation theory [34, 35], and coupled-cluster theory [35–37], have been used within the context of the supermolecular methods.

In the perturbational methods, the interaction energy can be computed without the need to determine the energy of the monomers or dimer. One commonly used method in this approach is symmetry-adapted perturbation theory (SAPT). In SAPT, the Hamiltonian of the dimer is written as a function of the contributions from each monomer and the perturbation. Compared to the supermolecular methods, the perturbational methods better preserve the nature of intermolecular interactions [28–30].

Within the molecular scale, atomistic molecular dynamics simulations may be impractical for describing systems with macromolecules (*e.g.*, polymers and proteins) or long-time phenomena (*e.g.*, self-assembly and slow diffusion processes). Instead, a coarse-graining methodology can be used to simplify the atomistic approach. In coarse-graining methods, a group of atoms is represented as a bead. The interactions between bonded and non-bonded beads can be determined to replicate specific properties of the studied system [27, 38]. Coarse-grained models are inserted within the meso scale, allowing larger time and size scales in molecular dynamics simulations with lower computational costs (see Fig. 1).

Reducing the detail of coarse-grained models implies describing matter at the macro scale. One of the few approaches in the literature that allows the connection between the molecular and the macro scale is the statistical associating fluid theory (SAFT) equations of state (see Fig. 1). SAFT equations of state are based on the first-order thermodynamic perturbation theory (TPT1) by Wertheim [39–43]. In SAFT, molecules are represented as segments of spheres that can form chains and ring structures. Each spherical segment has a repulsive core and multiple attractive sites for association. The total Helmholtz free energy of the system is described as the sum of the segment energy and a perturbation term. Variants of SAFT equations of state have been successfully used to study the thermodynamics of complex systems, ranging from simple hydrocarbons to electrolytes, polymers, pharmaceuticals, and confined fluids [1, 44, 45].

The SAFT family of equations of state parameters can be obtained using top-down and bottom-up methodologies. The reviews available in the literature [1, 44, 45] for variants of SAFT equations of state present countless references in which the parameters are fitted directly to experimental property data through the top-down methodology. Using the bottom-up methodology, the first attempt to obtain parameters of an equation of state through *ab initio* calculations was made by van Nhu *et al.* [4] through the perturbed-chain polar statistical associating fluid equation of state (PCP-SAFT).

In their work, van Nhu *et al.* [4] established the relationship of a series of molecular descriptors (*e.g.*, size, shape, charge distributions, and dispersion interaction of a molecule) with parameters of the PCP-SAFT equation of state. Good prediction results on experimental property data were only obtained when considering at least one of the experimental normal boiling point data parameters [4]. Although a drastic reduction of the experimental property data is promoted with the methodology of van Nhu *et al.* [4], a purely predictive character of the PCP-SAFT equation of state remains to be achieved. Most of the few works in the literature [5–9] are based on modifications to the methodology established by van Nhu *et al.* [4] using the PCP-SAFT equation of state.

Very recently, Walker *et al.* [10] proposed an alternative route for obtaining SAFT-VR Mie equation of state parameters through *ab initio* calculations. In their work, Walker *et al.* [10] fitted SAFT-VR Mie equation of state parameters to dimer interaction energy curves for some noble gases and methane. Although good prediction results on thermophysical properties and critical point were obtained by Walker *et al.* [10], the associated deviations are still large for applications in the design of chemical plants. Furthermore, within the same context investigated by Walker *et al.* [10], some of the results of this dissertation have already been published in a parallel and independent study [46], contributing to the investigation of systems composed of more complex molecules than noble gases and methane.

1.3 Objectives

The main objective of this dissertation is to develop a consistent methodology to obtain reliable interatomic potential parameters from *ab initio* calculations to give molecular-based, equations of state a purely predictive character.

This work also comprises secondary objectives such as:

- to obtain interatomic potential parameters from *ab initio* for a series of molecules of different chemical classes by supermolecular and perturbational methods;

- to compare the values of the interatomic potential parameters obtained by the top-down methodology available in the literature within the SAFT framework with the values of the parameters obtained in this study by the bottom-up methodology through *ab initio* calculations data;
- to compare the ability to predict thermophysical properties, critical point, and saturation properties when using top-down and bottom-up based interatomic potential parameters in SAFT equations of state (in this study, SAFT-VR Mie equation of state specifically);
- to compare the capacity to predict thermophysical properties using top-down and bottom-up based interatomic potential parameters SAFT equations of state (in this study, SAFT-VR Mie equation of state specifically) and in more robust simulation methods (in this study, molecular dynamics simulations);
- to establish a relationship between the quantum and macro scales in developing predictive models such as the SAFT equations of state.

2 Methodology

"If you think you understand quantum mechanics, you don't understand quantum mechanics."
Richard Feynman

2.1 Theoretical

2.1.1 Ab initio methods

At the quantum scale, studying the structure of molecules is closely related to understanding systems with many electrons and nuclei. In this context, analyzing the potential generated by the arrangement of electron distribution in nuclei allows for extracting information about the stability and properties of a studied system. As Classical Mechanics cannot correctly describe subatomic particles, there is no other approach to this type of problem than the theory of Quantum Mechanics.

A considerable part of Quantum Mechanics is grounded on the work of Schrödinger [19], who based his effort on the theories of de Broglie (wave-particle duality) and Hamilton-Jacobi (mathematical representation of a particle as a wave). The starting point for studying molecular systems in Quantum Mechanics is the Schrödinger's equation, which in its simplified and time-independent form is given by Eq. (2.1):

$$\hat{\mathscr{H}}\Psi(\mathbf{r}) = E\Psi(\mathbf{r}), \tag{2.1}$$

where $\hat{\mathscr{H}}$ is the Hamiltonian operator, Ψ is the wave function, E is the total energy, and $\mathbf{r}$ is the vector of spatial coordinates for electrons and nuclei.

For modeling a molecular system of "N" electrons and "M" nuclei in the absence of external fields, the Hamiltonian operator of the Schrödinger's equation can be calculated through Eq. (2.2) by five essential contributions given by Eqs. (2.3)-(2.7):

$$\hat{\mathscr{H}} = \hat{T}_{\mathrm{e}} + \hat{T}_{\mathrm{n}} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{nn}} + \hat{V}_{\mathrm{en}}, \tag{2.2}$$

with

$$\hat{T}_{\mathrm{e}} = -\sum_{i=1}^{N} \frac{\hbar}{2m_i}\nabla_i^2, \tag{2.3}$$

$$\hat{T}_{\mathrm{n}} = -\sum_{a=1}^{N} \frac{\hbar}{2m_a}\nabla_a^2, \tag{2.4}$$

$$\hat{V}_{\text{ee}} = \sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{e^2}{4\pi\varepsilon_0}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \tag{2.5}$$

$$\hat{V}_{\text{nn}} = \sum_{a=1}^{M}\sum_{b\neq a}^{M} \frac{Z_a Z_b e^2}{4\pi\varepsilon_0}\frac{1}{|\mathbf{r}_a - \mathbf{r}_b|}, \tag{2.6}$$

$$\hat{V}_{\text{en}} = \sum_{i=1}^{N}\sum_{a=1}^{M} \frac{Z_a e^2}{4\pi\varepsilon_0}\frac{1}{|\mathbf{r}_i - \mathbf{r}_a|}, \tag{2.7}$$

where $\hat{T}_{\text{e}}$ is the kinetic energy of the electrons, $\hat{T}_{\text{n}}$ is the kinetic energy of the nuclei, $\hat{V}_{\text{ee}}$ is the electron-electron repulsion energy, $\hat{V}_{\text{nn}}$ is the nucleus-nucleus repulsion energy, $\hat{V}_{\text{en}}$ is the electron-nucleus attraction energy, $\hbar$ is the reduced Planck constant, m is the mass, ∇^2 is the Laplacian operator, $\left(i.e., \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)$, e is the elementary charge, $\mathbf{r}$ is the vector of spatial coordinates, Z is the charge of the nucleus, ε_0 is the vacuum permittivity subscripts i and j are related to the electrons, and subscripts a and b are related to the nuclei.

The first simplification that can be done in Eq. (2.2) is the change to atomic units[1] with the elimination of most of the constants, resulting in Eq. (2.8):

$$\hat{\mathscr{H}} = \underbrace{-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2}_{\hat{T}_{\text{e}}} \underbrace{-\sum_{a=1}^{N}\frac{1}{2m_a}\nabla_a^2}_{\hat{T}_{\text{n}}} \underbrace{+\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\hat{V}_{\text{ee}}} \underbrace{+\sum_{a=1}^{M}\sum_{b\neq a}^{M}\frac{Z_a Z_b}{|\mathbf{r}_a - \mathbf{r}_b|}}_{\hat{V}_{\text{nn}}} \underbrace{-\sum_{i=1}^{N}\sum_{a=1}^{M}\frac{Z_a}{|\mathbf{r}_i - \mathbf{r}_a|}}_{\hat{V}_{\text{en}}}, \tag{2.8}$$

The problem of many electrons and nuclei systems can be further simplified applying the Born-Oppenheimer approximation. Within the Born-Oppenheimer approximation, nuclei can be treated as static since nuclei are much heavier than electrons. Consequently, the motion of electrons can be treated independently of the nuclear motion. Thus, the electronic component of the Hamiltonian, which represents the movement of "N" electrons in the field of "M" fixed nuclei, can be separated and obtained summing three contributions given in atomic units by Eq. (2.9):

$$\hat{\mathscr{H}}_{\text{elec}} = \underbrace{-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2}_{\hat{T}_{\text{e}}} \underbrace{+\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\hat{V}_{\text{ee}}} \underbrace{-\sum_{i=1}^{N}\sum_{a=1}^{M}\frac{Z_a}{|\mathbf{r}_i - \mathbf{r}_a|}}_{\hat{V}_{\text{en}}}. \tag{2.9}$$

For most applications in molecular systems, the aim is the solution of the nonrelativistic, time-independent, and fixed-nuclei Schrödinger's equation, given by Eq. (2.10):

$$\hat{\mathscr{H}}_{\text{elec}}\Psi_{\text{elec}}(\mathbf{r}) = E_{\text{elec}}\Psi_{\text{elec}}(\mathbf{r}), \tag{2.10}$$

wherein $\hat{\mathscr{H}}_{\text{elec}}$ is the electronic Hamiltonian operator, Ψ_{elec} is the electronic wave function, E_{elec} is the electronic energy, and $\mathbf{r}$ is the vector of spatial coordinates for electrons or nuclei.

[1] In atomic units, mass is given in units of the electron mass, charge is given in units of elementary charge, and length is given in units of the Bohr radius.

The electronic wave function presented in Eq. (2.10) may differ for different arrangements of nuclei. Consequently, the total energy for systems with "M"-fixed nuclei is also composed of a constant given by Eq. (2.11) that represents the nuclear repulsion:

$$E = E_{\text{elec}} + \underbrace{\sum_{a=1}^{M} \sum_{b \neq a}^{M} \frac{Z_a Z_b}{|\mathbf{r}_a - \mathbf{r}_b|}}_{\hat{V}_{\text{nn}}}. \tag{2.11}$$

Even decoupling nuclear and electronic motion via the Born-Oppenheimer approximation, the dynamics of systems of many electrons are still too complex to deal with them analytically. Schrödinger's equation lacks an analytical solution for systems with more than one electron (even atoms or molecules). As a result, the development of computational methods that allow the generation of approximate numerical solutions for this type of problem is desired.

The methods to address the many-electron problem that generate solutions without the need for reference to experimental data are called *ab initio*. Some relevant existing *ab initio* methods are Hartree-Fock, post-Hartree-Fock ones, and density functional theory.

2.1.2 Hartree-Fock method

Hartree-Fock is the simplest *ab initio* method. The Hartree-Fock method is based on the works of Hartree [20], Slater [21, 22], and Fock [23, 24]. In 1928, Hartree [20] introduced a method called the self-consistent field for computing the wave function for many electron systems. In this method, Hartree [20] considered the electrons as uncorrelated (*i.e.*, independent of each other) to construct the electronic wave function.[2] Thus, the wave function for "N" uncorrelated electrons can be written as a product of "N" wave functions of one electron, and the resulting Eq. (2.12) is called the Hartree product:

$$\Psi(\mathbf{r}) = \Psi(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = \psi_1(\mathbf{r}_1)\,\psi_2(\mathbf{r}_2)\cdots\psi_N(\mathbf{r}_N) = \prod_{i=1}^{N} \psi_i(\mathbf{r}_i), \tag{2.12}$$

where Ψ is the electronic wave function, ψ is the one-electron wave function, $\mathbf{r}$ is the vector of spatial coordinates, and subscripts 1, 2, $\cdots$ N are labels for electrons.

The Eq. (2.12) for the Hartree product has two main issues. The first one is related to a quantum property of electrons called spin. The spin describes the direction of rotation of an electron in the presence of a magnetic field, making possible the existence of states of alignment along or opposite the field. Therefore, to describe the electron completely, in addition to the spatial coordinates, it is also necessary to specify the spin.

[2] From now on, only the electronic problem will be considered, with the subscript elec being suppressed from the equations.

For this purpose, the one-electron wave function (*i.e.*, one-electron orbital) must be rewritten as a product between a component that represents the spatial coordinates (*i.e.*, the spatial orbital) and another contribution that represents the spin. The spin contribution is described by two functions that must be orthonormal. The one-electron wave functions (*i.e.*, one-electron orbitals), now called spin orbitals, are functions of three spatial coordinates and a coordinate referring to the spin quantum property as described by Eq. (2.13):

$$\chi_i(\mathbf{x}_i) = \begin{cases} \psi_i(\mathbf{r}_i)\,\alpha(\mathbf{s}) \\ \psi_i(\mathbf{r}_i)\,\beta(\mathbf{s}) \end{cases}, \quad [3] \tag{2.13}$$

where χ is the spin orbital, ψ is the spatial orbital (*i.e.*, one-electron wave function), α represents the spin-up state ($\uparrow$), β represents the spin-down state ($\downarrow$), $\mathbf{s}$ is the vector that represents the spin coordinate, $\mathbf{x}$ is a vector that represents three spatial coordinates and one spin coordinate (*i.e.*, $\mathbf{x} = \{\mathbf{r}, \mathbf{s}\}$), and subscript i is a label for electronic states.

The second problem associated with Eq. (2.12) for the Hartree product is that it is inconsistent with the indistinguishability of electrons. By the strictest statement of the Pauli exclusion principle, the total wave function of a many-particle system must be antisymmetric. As an implication, the exchange of spatial coordinates and spin of two electrons must change the sign of the total wave function, according to Eq. (2.14):

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N) = -\Psi(\mathbf{x}_2, \mathbf{x}_1, \cdots, \mathbf{x}_N). \tag{2.14}$$

As a consequence of Eq. (2.14), the wavefunction for "N" uncorrelated electrons cannot be written as a product of "N" wavefunctions of one electron, as demonstrated by Eq. (2.15):

$$\chi_1(\mathbf{x}_1)\,\chi_2(\mathbf{x}_2)\cdots\chi_N(\mathbf{x}_N) \neq \chi_1(\mathbf{x}_2)\,\chi_2(\mathbf{x}_1)\cdots\chi_N(\mathbf{x}_N). \tag{2.15}$$

The verification that the Hartree product does not obey the principle of antisymmetry of the wave function was independently pointed out by Slater [21, 22] and Fock [23, 24] in 1930. As a solution to respect the Pauli exclusion principle, the Hartree-Fock method uses a single Slater determinant as a mathematical device to antisymmetrize the wave function, resulting in

[3] Since the α and β functions to characterize the spin of electrons are orthonormal, they satisfy the following conditions: $\int \alpha^*(\mathbf{s})\,\alpha(\mathbf{s})\,d\mathbf{s} = \int \beta^*(\mathbf{s})\,\beta(\mathbf{s})\,d\mathbf{s} = 1$ and $\int \alpha^*(\mathbf{s})\,\beta(\mathbf{s})\,d\mathbf{s} = \int \beta^*(\mathbf{s})\,\alpha(\mathbf{s})\,d\mathbf{s} = 0$, where superscript $*$ represents the complex conjugate of the function. In Dirac or bra-ket notation for integrals these previous equations can be written as $\langle\alpha|\alpha\rangle = \langle\beta|\beta\rangle = 1$ and $\langle\alpha|\beta\rangle = \langle\beta|\alpha\rangle = 0$.

Eq. (2.16), known as the approximation of Hartree-Fock:

$$\begin{aligned}\Psi(\mathbf{x}) = \Psi(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N) &= \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \vdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \\ &= \frac{1}{\sqrt{N!}} \sum_{n=1}^{N!} (-1)^{p_n} \hat{P}_n \{\chi_1(1)\chi_2(2)\cdots\chi_N(N)\} \\ &= \hat{\mathscr{A}} \{\chi_1(1)\chi_2(2)\cdots\chi_N(N)\},\end{aligned} \tag{2.16}$$

where p_n is 0 or 1 if n is even or odd, respectively; $\hat{P}_n$ is an permutation operator that generates the n-th permutation at the indices of $\mathbf{x}$, $\hat{\mathscr{A}}$ is the anti-symmetrization operator, and subscripts 1, 2, $\cdots$ N are labels for electronic states.

Since the wave function of many electrons is approximated by a single Slater determinant, the electron correlation effects cannot be fully explained in the Hartree-Fock method. The essence of the Hartree-Fock approximation is to treat the electron-electron repulsion on average. In the Hartree-Fock method, the electrons are moving at a mean spherical potential due to the other electrons and nuclei of the atom or molecule.

The problem to be solved in the Hartree-Fock method is to find spin orbitals to minimize the true energy (ground state) of the system. In this context, the variational principle becomes a useful mathematical strategy to carry out the wave function optimization process (since the wave function is dependent on the spin orbitals). According to the variational principle, the computed energy using a trial electronic wave function will then be greater than or equal to the lowest true energy (ground-state) of the system, according to Eq. (2.17) in Dirac or bra-ket notation for integrals:

$$E = \frac{\langle \Psi_{\text{trial}} | \hat{\mathscr{H}} | \Psi_{\text{trial}} \rangle}{\langle \Psi_{\text{trial}} | \Psi_{\text{trial}} \rangle} \geqslant E_0,^{4} \tag{2.17}$$

where E is the electronic energy, Ψ_{trial} is a trial electronic wave function, $\hat{\mathscr{H}}$ is the electronic Hamiltonian, and E_0 is true electronic energy (ground-state).

By the condition of orthonormality of the spin orbitals (see Eq. (2.13) footnotes for further details), the problem to be solved is to minimize the electronic energy functional[5] given

[4] In Eq. (2.17):
$\langle \Psi_{\text{trial}} | \hat{\mathscr{H}} | \Psi_{\text{trial}} \rangle = \int \Psi^*_{\text{trial}}(\mathbf{x}) \hat{\mathscr{H}} \Psi_{\text{trial}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}$, and
$\langle \Psi_{\text{trial}} | \Psi_{\text{trial}} \rangle = \int \Psi^*_{\text{trial}}(\mathbf{x}) \Psi_{\text{trial}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}$,
wherein superscript $*$ represents the complex conjugate of the function.

[5] While a function is a rule that transforms a number into another number $\left(e.g., 2 \xrightarrow{x^3} 8\right)$, a functional is a rule that transforms a function into a number $\left(e.g., x^3 \xrightarrow{\int_0^2 x^3 dx} \frac{x^4}{4}\Big|_0^2 = 4\right)$.

by Eq. (2.18) in Dirac or bra-ket notation for integrals:

$$E\left[\chi\left(\mathbf{x}\right)\right] = \left\langle \Psi | \hat{\mathscr{H}} | \Psi \right\rangle .^{6} \tag{2.18}$$

The energy calculation of Eq. (2.18) can be simplified by separating the electronic Hamiltonian of Eq. (2.8) into one-electron and two-electron operators by Eqs. (2.19) and (2.20), respectively:

$$\hat{h}_1\left(i\right) = -\frac{1}{2}\nabla_i^2 - \sum_{a=1}^{M}\frac{Z_a}{\left|\mathbf{r}_i - \mathbf{r}_a\right|}, \tag{2.19}$$

$$\hat{h}_2\left(i,j\right) = \frac{1}{\left|\mathbf{r}_i - \mathbf{r}_j\right|}, \tag{2.20}$$

where $\hat{h}_1$ is the one-electron Hamiltonian operator, and $\hat{h}_2$ is the two-electron Hamiltonian operator.

Consequently, with the definitions of Eqs. (2.19) and (2.20) the electronic Hamiltonian operator of Eq. (2.8) can be written in a more compact form as shown in Eq. (2.21):

$$\hat{\mathscr{H}} = \sum_{i=1}^{N}\hat{h}_1\left(i\right) + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\hat{h}_2\left(i,j\right). \tag{2.21}$$

Thus, the electronic energy functional of Eq. (2.18) can now be rewritten in terms of one- and two-electron integrals, according to Eq. (2.22) in Dirac Notation or bra-ket for integrals:

$$\begin{aligned} E\left[\chi\left(\mathbf{x}\right)\right] =& \left\langle \chi_1\chi_2\cdots\chi_N \right| \sum_{i=1}^{N}\hat{h}_1\left(i\right) + \sum_{i=1}^{N}\sum_{j\neq i}^{N}\hat{h}_2\left(i,j\right)\left|\chi_1\chi_2\cdots\chi_N\right\rangle, \\ =& \sum_{i=1}^{N}\underbrace{\left\langle \chi_i\right|\hat{h}_1\left(i\right)\left|\chi_i\right\rangle}_{h_i} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\Big[\underbrace{\left\langle \chi_i\chi_j\right|\hat{h}_2\left(i,j\right)\left|\chi_i\chi_j\right\rangle}_{J_{ij}} - \underbrace{\left\langle \chi_i\chi_j\right|\hat{h}_2\left(i,j\right)\left|\chi_j\chi_i\right\rangle}_{K_{ij}}\Big], \end{aligned}^{7} \tag{2.22}$$

---

[6] In Eq. (2.18): $\left\langle \Psi | \hat{\mathscr{H}} | \Psi \right\rangle = \int \Psi^{*}\left(\mathbf{x}\right)\hat{\mathscr{H}}\Psi\left(\mathbf{x}\right)\mathrm{d}\mathbf{x}$, where superscript $*$ represents the complex conjugate of the function.

[7] In Eq. (2.22):
$\left\langle \chi_i\right|\hat{h}_1\left(i\right)\left|\chi_i\right\rangle = \int \chi_i^{*}\left(\mathbf{x}_1\right)\hat{h}_1\left(\mathbf{r}_1\right)\chi_i\left(\mathbf{x}_1\right)\mathrm{d}\mathbf{x}_1$,
$\left\langle \chi_i\chi_j\right|\hat{h}_2\left(i,j\right)\left|\chi_i\chi_j\right\rangle = \int \chi_i^{*}\left(\mathbf{x}_1\right)\chi_j^{*}\left(\mathbf{x}_2\right)r_{12}^{-1}\chi_i\left(\mathbf{x}_1\right)\chi_j\left(\mathbf{x}_2\right)\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2 = J_{ij}$, and
$\left\langle \chi_i\chi_j\right|\hat{h}_2\left(i,j\right)\left|\chi_j\chi_i\right\rangle = \int \chi_i^{*}\left(\mathbf{x}_1\right)\chi_j^{*}\left(\mathbf{x}_2\right)r_{12}^{-1}\chi_j\left(\mathbf{x}_1\right)\chi_i\left(\mathbf{x}_2\right)\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2 = K_{ij}$,
where $\mathbf{x}_1$ and $\mathbf{x}_2$ are spin coordinates for a pair formed by electrons 1 and 2, respectively; r_{12}^{-1} is the distance between two electrons for a pair formed by electrons 1 and 2, respectively, and superscript $*$ represents the complex conjugate of the function.
Furthermore, in related literature it is common to find Eq. (2.22) written in a more compact form as
$E\left[\chi\left(\mathbf{x}\right)\right] = \sum_{i=1}^{N}\left\langle i|\hat{h}|i\right\rangle + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left\langle ij||ij\right\rangle$ or $E\left[\chi\left(\mathbf{x}\right)\right] = \sum_{i=1}^{N}\left\langle i|\hat{h}|i\right\rangle + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left([ii|jj] - [ij|ji]\right)$,
with one-electron integral given by
$\left\langle i|\hat{h}|i\right\rangle = \int \chi_i^{*}\left(\mathbf{x}_1\right)\hat{h}_1\left(\mathbf{r}_1\right)\chi_i\left(\mathbf{x}_1\right)\mathrm{d}\mathbf{x}_1$,
and two-electron integral (in Chemists' notation) given by
$[ij|kl] = \langle ik|jl\rangle = \int \chi_i^{*}\left(\mathbf{x}_1\right)\chi_j\left(\mathbf{x}_2\right)r_{12}^{-1}\chi_k^{*}\left(\mathbf{x}_1\right)\chi_l\left(\mathbf{x}_2\right)\mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2$.

where h_i is the one-eletron integral, J_{ij} is the Coulomb integral (represents the classical Coulomb repulsion), and K_{ij} is the exchange integral (represents the manifestation of the correlation between the movements of electrons).

In the Hartree-Fock method, the task is reduced to minimizing the electronic energy functional given Eq. (2.22) subject to the constraint given by Eq.(2.23) in Dirac or bra-ket notation for integrals, which guarantees the orthonormality of the spin-orbitals:

$$\left\langle \chi_i | \chi_j \right\rangle = \delta_{ij},^{8}. \tag{2.23}$$

where δ_{ij} is a Kronecker delta function (*i.e.*, $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$).

The method of undetermined Lagrange multipliers can be used to solve this problem, which involves optimizing the electronic wave function to minimize the energy functional. Through the method of undetermined Lagrange multipliers, the objective is to minimize the Lagrange functional given by Eq. (2.25):

$$\mathscr{L}\left[\chi\left(\mathbf{x}\right)\right] = E\left[\chi\left(\mathbf{x}\right)\right] - \sum_{i=1}^{N}\sum_{j\neq i}^{N} \varepsilon_{ij}\left(\left\langle \chi_i | \chi_j \right\rangle - \delta_{ij}\right), \tag{2.24}$$

where $\mathscr{L}$ is the Lagrangian operator, and ε_{ij} are the Lagrange multipliers (represent the orbital energy).

The minimum of the electronic energy functional occurs when the variation of the Lagrange functional is equal to zero $(i.e., \delta\mathscr{L} = 0)$. This can be accomplished by applying an infinitesimal variation in the spin orbitals $\left(i.e., \chi_i \rightarrow \chi_i + \delta\chi_j\right)$. Thus, for the electronic energy functional to be minimal, it is necessary to solve Eq. (2.25):

$$\delta\mathscr{L}\left[\chi\left(\mathbf{x}\right)\right] = \delta E\left[\chi\left(\mathbf{x}\right)\right] - \sum_{i=1}^{N}\sum_{j\neq i}^{N} \varepsilon_{ij}\delta\left\langle \chi_i | \chi_j \right\rangle = 0. \tag{2.25}$$

After performing some algebraic manipulations on Eq. (2.25) (whose steps can be readily found in the literature), the set of N-coupled equations of the Hartree-Fock method can be obtained, according to Eq. (2.26):

$$\underbrace{\left\{\hat{h}_1\left(\mathbf{r}_1\right) + \overbrace{\sum_{j\neq i}^{N}\left[\hat{\mathscr{J}}_j\left(\mathbf{x}_1\right) - \hat{\mathscr{K}}_j\left(\mathbf{x}_1\right)\right]}^{\hat{v}^{\mathrm{HF}}\left(\mathbf{x}_1\right)}\right\}}_{\hat{\mathscr{F}}\left(\mathbf{x}_1\right)} \chi_i\left(\mathbf{x}_1\right) = \sum_{j\neq i}^{N} \varepsilon_{ij}\chi_j\left(\mathbf{x}_1\right) \tag{2.26}$$
$$\text{for } i = 1, 2, \cdots, N;^{9}$$

[8] In Eq. (2.23):
$\left\langle \chi_i | \chi_j \right\rangle = \int \chi_i^*\left(\mathbf{x}_1\right)\chi_j\left(\mathbf{x}_1\right)\mathrm{d}\mathbf{x}_1,$
where $\mathbf{x}_1$ are spin coordinates for a single electron and superscript $*$ represents the complex conjugate of the function.

[9] In Eq. (2.26):

where $\hat{\mathscr{F}}$ is called Fock operator, $\hat{v}^{\mathrm{HF}}$ is the Hartree-Fock effective potential operator, $\hat{\mathscr{J}}$ is called Coulomb operator, $\hat{\mathscr{K}}$ is called exchange operator, $\mathbf{r}_1$ are spatial coordinates for a single electron. and $\mathbf{x}_1$ are spin coordinates for a single electron.

The previously introduced N-coupled equations of the Hartree-Fock are not in their canonical form. To achieve this, a unitary transformation in the spin orbitals, as specified in Eq. (2.27), is required:

$$\chi_i^{'} = \sum_{j \neq i}^{N} \chi_j \hat{\mathscr{U}}_{ij}, \tag{2.27}$$

where $\chi^{'}$ are spin orbitals that guarantee the matrix with elements formed by Lagrange multipliers (*i.e.*, ε_{ij}) is diagonal, and $\hat{\mathscr{U}}$ is the unitary transformation operator.

By substituting Eq. (2.26) into Eq. (2.25), the Hartree Fock equation in its canonical form given by Eq. (2.28) is obtained:

$$\begin{gathered} \hat{\mathscr{F}}(\mathbf{x}_1)\chi_i^{'}(\mathbf{x}_1) = \varepsilon_i^{'}\chi_i^{'}(\mathbf{x}_1) \\ \text{for } i = 1, 2, \cdots, N; \end{gathered} \tag{2.28}$$

where $\varepsilon^{'}$ is the energy associated with the canonical spin orbitals.

The Hartree Fock equation in its canonical form given by Eq. (2.28) can be solved numerically (*i.e.*, exact Hartree-Fock) or by expanding the spin orbitals through a set of functions known as basis functions, as described by Eq. (2.29), in a method independently proposed by Roothaan [47] and Hall [48] in 1951 (*i.e.*, Hartree-Fock-Roothan equations):

$$\begin{gathered} \chi_i = \sum_{\mu=1}^{K} C_{\mu i}\phi_\mu, \\ \text{for } i = 1, 2, \cdots, N; \end{gathered} \tag{2.29}$$

where K is the number of basis functions, C are expansion coefficients to be determined, and ϕ are basis functions.

In addition, the Hartree-Fock approach has three variants to treat the function that represents the spin in the spin orbitals: restricted closed-shell, restricted open-shell, and unrestricted open-shell. In restricted variants of the Hartree-Fock approach, atoms or molecules have an even number of electrons, so the orbitals are doubly occupied with electrons of opposite spin. The difference between the restricted closed and open-shell variants is that the valence shell is not completely filled with electrons in the open shell. In the unrestricted open-shell variant of the

$\hat{\mathscr{J}}_j(\mathbf{x}_1)\chi_i(\mathbf{x}_1) = \left[\int \chi_j^*(\mathbf{x}_2) r_{12}^{-1} \chi_j(\mathbf{x}_2)\, \mathrm{d}\mathbf{x}_2\right]\chi_i(\mathbf{x}_1) = \left[\int |\chi_j(\mathbf{x}_2)|^2 r_{12}^{-1}\, \mathrm{d}\mathbf{x}_2\right]\chi_i(\mathbf{x}_1)$, and

$\hat{\mathscr{K}}_j(\mathbf{x}_1)\chi_i(\mathbf{x}_1) = \left[\int \chi_j^*(\mathbf{x}_2) r_{12}^{-1} \chi_i(\mathbf{x}_2)\, \mathrm{d}\mathbf{x}_2\right]\chi_j(\mathbf{x}_1)$,

where r_{12}^{-1} is the distance between two electrons 1 and 2 in different spin orbitals, superscript $*$ represents the complex conjugate of the function.

Hartree-Fock approach, there is an odd number of electrons, and the energy must be calculated considering all spins of orbitals occupied by a single electron.

For the restricted variant of the Hartree-Fock approach, if the definition of basis functions given by Eq. (2.29) is replaced in the Hartree Fock equation in its canonical form given by Eq. (2.28), the result obtained is Eq. (2.30):

$$\begin{gathered}\hat{\mathscr{F}}(\mathbf{x}_i)\underbrace{\sum_{\nu=1}^{K} C_{\nu i}\phi_\nu}_{\chi_i(\mathbf{x}_1)} = \varepsilon_i \underbrace{\sum_{\nu=1}^{K} C_{\nu i}\phi_\nu}_{\chi_i(\mathbf{x}_1)} \\ \text{for } i = 1,2,\cdots,N;\end{gathered} \tag{2.30}$$

By multiplying Eq. (2.30) on the left by a complex conjugate of the basis function (*e.g.*, ϕ_μ^*) and integrating, the problem can be transformed into matrix form, which is more computer-friendly, according to Eq. (2.31) in Dirac or bra-ket notation for integrals:

$$\begin{gathered}\sum_{\nu=1}^{K} C_{\nu i}\underbrace{\left\langle \phi_\mu \middle| \hat{\mathscr{F}}(\mathbf{x}_1) \middle| \phi_\nu \right\rangle}_{F_{\mu\nu}} = \varepsilon_i \sum_{\nu=1}^{K} C_{\nu i}\underbrace{\left\langle \phi_\mu \middle| \phi_\nu \right\rangle}_{S_{\mu\nu}},{}^{10} \\ \text{for } i = 1,2,\cdots,N;\end{gathered} \tag{2.31}$$

The expression in Eq. (2.31) can be further simplified, leading to the derivation of the Roothaan equations according to Eq. (2.32):

$$\mathbf{FC} = \mathbf{SC}\varepsilon \tag{2.32}$$

where $\mathbf{F}$ is the Fock matrix with elements given by $F_{\mu\nu} = \left\langle \phi_\mu \middle| \hat{\mathscr{F}}(\mathbf{x}_1) \middle| \phi_\nu \right\rangle$, $\mathbf{C}$ is the expansion coefficients matrix, $\mathbf{S}$ is the overlap matrix with elements given by $S_{\mu\nu} = \left\langle \phi_\mu \middle| \phi_\nu \right\rangle$, and ε is a diagonal matrix of the orbital energies.

In the context of the unrestricted variant of the Hartree-Fock method, the outcome consists of a pair of coupled Roothaan equations referred to as the Pople-Nesbet-Berthier equations, as presented in Eq. (2.33):

$$\begin{aligned}\mathbf{F}^\alpha\mathbf{C}^\alpha &= \mathbf{SC}^\alpha\varepsilon^\alpha \\ \mathbf{F}^\beta\mathbf{C}^\beta &= \mathbf{SC}^\beta\varepsilon^\beta\end{aligned} \tag{2.33}$$

where α represents the spin-up state ($\uparrow$), and β represents the spin-down state ($\downarrow$).

As in Eqs. (2.32) and (2.33) the Fock matrix depends on its own solution through the spin orbitals, an initial guess of the spin orbitals must be refined interactively up to a certain convergence criterion in a methodology called self-consistent field approach.

[10] In Eq. (2.29):
$\left\langle \phi_\mu \middle| \hat{\mathscr{F}}(\mathbf{x}_1) \middle| \phi_\nu \right\rangle = \int \phi_\mu^*(\mathbf{x}_1)\,\hat{\mathscr{F}}(\mathbf{x}_1)\,\phi_\nu(\mathbf{x}_1)\,\mathrm{d}\mathbf{x}_1 = F_{\mu\nu}$, and
$\left\langle \phi_\mu \middle| \phi_\nu \right\rangle = \int \phi_\mu^*(\mathbf{x}_1)\,\phi_\nu(\mathbf{x}_1)\,\mathrm{d}\mathbf{x}_1 = S_{\mu\nu}$.

2.1.3 Post Hartree-Fock methods

Originally, the Hartree-Fock method assumes that an electron is subject to a mean spherical potential due to the other electrons and nuclei in an atom and a molecule. As the electron-electron repulsion is treated in an average way, the details of the interactions of each pair of electrons are disregarded. Even though the wave function is described as precise as possible, there is an error in the total energy obtained by the Hartree-Fock method.

One approach to correct the total Hartree-Fock energy is to consider an electronic correlation, which is given by the difference between the true energy (ground-state) and the Hartree-Fock energy, according to Eq. (2.34):

$$E_{\text{corr}} = E_0 - E^{\text{HF}}, \tag{2.34}$$

where E^{HF} is the Hartree-Fock energy.

The methods in which the correlation energy is determined are called post-Hartree-Fock. Two post-Hartre-Fock methods widely discussed in the literature are the Møller-Plesset perturbation theory and the coupled cluster approximation.

2.1.3.1 Møller-Plesset perturbation theory

The Møller-Plesset approach for dealing with electronic correlation and improving the results of the Hartree-Fock method emerged with the work of Møller and Plesset [49] in 1934, and is based on the application of second-order perturbation theory. The idea of the Møller-Plesset perturbation theory is that a complex system (closer to reality) can be treated as a perturbed version of a simpler system (idealized).

According to the Rayleigh-Schrödinger's perturbation theory, the Hamiltonian can be written as a sum of the unperturbed Hamiltonian and a perturbation, according to Eq. (2.35):

$$\hat{\mathscr{H}} = \hat{\mathscr{H}}_0 + \lambda \hat{\mathscr{V}}, \tag{2.35}$$

where $\hat{\mathscr{H}}$ is the Hamiltonian, $\hat{\mathscr{H}}_0$ is the unperturbed Hamiltonian, λ is a coupling parameter (a dimensionless parameter that defines the order of perturbation), and $\hat{\mathscr{V}}$ is the perturbation operator.

For an "N" electron system, in the the Møller-Plesset perturbation theory the unperturbed Hamiltonian and the perturbation are described by Eqs. (2.36) and (2.37), respectively:

$$\hat{\mathscr{H}}_0 = \sum_{i=1}^{N} \hat{\mathscr{F}}(i) = \sum_{i=1}^{N} \left[\hat{h}_1(i) + \hat{v}^{\text{HF}}(i)\right], \tag{2.36}$$

$$\hat{\mathscr{V}} = \sum_{i=1}^{N} \left[\sum_{j \neq i}^{N} \frac{1}{\left|\mathbf{r}_i - \mathbf{r}_j\right|} - \hat{v}^{\text{HF}}(i) \right]. \tag{2.37}$$

In the Møller-Plesset perturbation theory, there is a hierarchy of energy levels in which electron-electron repulsion is treated in more detail. To identify the hierarchy, the abbreviation "MPn" is used, where "n" reflects the degree of the perturbation considered (*e.g.*, MP0, MP1, MP2, and so on). The energy level MP0 considers only the sum of the contributions of the energies of a one-electron operator given by Hartree-Fock and works as an unperturbed operator. The energy level MP1 corresponds to the energy level MP0 corrected with the Coulomb and exchange integrals (*i.e.*, MP1 refers to the Hartree-Fock energy), according to Eq. (2.38):

$$E^{\mathrm{MP1}} = E^{\mathrm{HF}} = E_0^{(0)} + E_0^{(1)} = E^{\mathrm{MP0}} + E_0^{(1)} = \langle \Psi_0^{(0)} | \hat{\mathscr{H}}_0 + \hat{\mathscr{V}} | \Psi_0^{(0)} \rangle, \tag{2.38}$$

where superscripts (0) and (1) are related to zeroth and first order perturbations, respectively, subscripts MP0 and MP1 are related to zeroth and first order Møller-Plesset perturbations, respectively, and subscript HF are related to Hartree-Fock.

The MP2 energy level represents the Hartree-Fock energy plus a correction term that describes the energy reduction due to repulsion between electrons rather than the Hartree-Fock treatment (*i.e.*, MP2 is the first energy level that goes beyond the Hartree-Fock treatment), according to Eq. (2.39):

$$E^{\mathrm{MP2}} = E^{\mathrm{HF}} + E_0^{(2)} = E^{\mathrm{HF}} + \sum_{k \neq 0}^{\infty} \frac{\left| \langle \Psi_0^{(0)} | \hat{\mathscr{V}} | \Psi_k^{(0)} \rangle \right|^2}{E_0^{(0)} - E_k^{(0)}}, \tag{2.39}$$

where subscript MP2 stands to the second order Møller-Plesset perturbation.

2.1.3.2 Coupled cluster approximation

The coupled cluster approximation arose to calculate the electronic correlation in atoms and molecules with Čížek [50] in 1966 and later with Paldus [51] in 1972. The coupled cluster approximation method is based on the linked cluster theorem, and the main idea is to treat a system with many electrons as clusters of few electrons. In this *ab initio* method, the interactions between the electrons of the same group are calculated, and, later, the interactions of electrons of different groups.

In the coupled cluster approximation method, the exact electronic wave function is related to the Hartree-Fock electronic wave function through a series of operators, according to Eq. (2.40):

$$\Psi = \underbrace{\left(1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \cdots\right)}_{e^{\hat{T}}} \Psi^{\mathrm{HF}}, \tag{2.40}$$

where Ψ is the electronic wave function given by the Slater determinant, $\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots$ with $\hat{T}_1, \hat{T}_2, \cdots$ as excitation operators for promoting one, two, $\cdots$ electrons into virtual spin orbitals.

By the amount of excitation operator terms (*i.e.*, $\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots$), the methods such as couple cluster doubles (CCD), couple cluster singles and doubles (CCSD) or couple cluster singles, doubles and triples (CCSDT) can be obtained, according to the set of Eqs. (2.41)-(2.43), respectively.

$$\hat{T}_{\mathrm{CCD}} = e^{\hat{T}_2}\Psi_{\mathrm{HF}}, \tag{2.41}$$

$$\hat{T}_{\mathrm{CCSD}} = e^{\left(\hat{T}_1+\hat{T}_2\right)}\Psi_{\mathrm{HF}}, \tag{2.42}$$

$$\hat{T}_{\mathrm{CCSDT}} = e^{\left(\hat{T}_1+\hat{T}_2+\hat{T}_3\right)}\Psi_{\mathrm{HF}}. \tag{2.43}$$

There is still a variant in which the contributions of the triple excitations are represented approximately, in which the method is called coupled-cluster with singles, doubles, and perturbative triples excitations (CCSD(T)), with a lower computational demand than the couple cluster singles, doubles and triples (CCSDT) method.

2.1.4 Density functional theory

Using electron density is an alternative route that allows solving the problem of many electrons without directly computing the electronic wave function. Formally, in an "N"-electron system, the electron density represents the probability of finding any one of the "N" electrons of any spin in a given volume element, where the positions and spins of the remaining "N-1" electrons are arbitrary.

The mathematical description for the electron density in an "N"-electron system is given by Eq. (2.44):

$$\rho\left(\mathbf{r}\right) = N\int\cdots\int\left|\Psi\left(\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_N\right)\right|^2 \mathrm{d}\mathbf{x}_1\mathrm{d}\mathbf{x}_2\cdots\mathrm{d}\mathbf{x}_N, \tag{2.44}$$

and from Eq. (2.44), the number of electrons in an atom or molecule can be obtained by integrating the electron density over the entire volume according to Eq. (2.46):

$$\int_V \rho\left(\mathbf{r}\right) d\mathbf{r} = N,^{11} \tag{2.45}$$

where ρ is the electron density, N is the number of electrons, Ψ is the electronic wave function, V is a volume element, $\mathbf{r}$ is the vector of spatial coordinates, $\mathbf{x}$ is a vector that represents three spatial coordinates and one spin coordinate (*i.e.*, $\mathbf{x} = \{\mathbf{r}, \mathbf{s}\}$), and subscripts 1, 2, $\cdots$ N are labels for electronic states.

Ab initio methods based on the calculation of electron density have two main advantages compared to those based on the electronic wave function. Unlike the electronic wave function, the first is that electron density is an experimentally observable quantity (*i.e.*, it can be verified by X-ray diffraction or electron diffraction). The second is that, while the wave

[11] In Eq (2.46), $\rho\left(\mathbf{r}\right) \geqslant 0$ for all $\mathbf{r}$ since a negative number of electrons per unit volume is impossible.

function of an "N"-electron system is a function of "$4N$" variables (*i.e.*, it considers three spatial coordinates and one spin coordinate times the number of electrons), the electron density is dependent on only three variables (*i.e.*, it considers three spatial coordinates).

The main objective of *ab initio* methods based on the calculation of electron density is to connect the electronic energy of the system with the electron density through functionals, as described by Eq (2.46):

$$E = F\left[\rho\right], \tag{2.46}$$

where F represents a generic functional.

The first attempt of a model that used the electron density rather than the wave function was designed independently by Thomas [52] and Fermi [53] in 1927. In the Thomas-Fermi model, where the system can be described as a uniform electron gas, the electronic energy is described by the functional of Eq. (2.47):

$$E^{\mathrm{TF}}\left[\rho\right] = T_{\mathrm{e}}\left[\rho\right] + V_{\mathrm{ee}}\left[\rho\right] + V_{\mathrm{en}}\left[\rho\right], \tag{2.47}$$

wherein E^{TF} is the approximate electronic energy in the Thomas-Fermi model, T_{e} is the kinetic energy for a uniform electron gas model, V_{ee} is the electron-electron repulsion containing only the Coulomb potential, and V_{en} is the electron-nucleus attraction energy.

An improvement to the original functional of the Thomas-Fermi model was proposed by Dirac [54] in 1930. In the Thomas-Fermi-Dirac model, the electron-electron repulsion is split into two contributions: a Coulombic term and an exchange expression. Thus, in the Thomas-Fermi-Dirac model, electronic energy is described by the functional of Eq. (2.48):

$$E^{\mathrm{TFD}}\left[\rho\right] = T_{\mathrm{e}}\left[\rho\right] + \underbrace{J\left[\rho\right] + K\left[\rho\right]}_{V_{\mathrm{ee}}\left[\rho\right]} + V_{\mathrm{en}}\left[\rho\right], \tag{2.48}$$

where E^{TFD} is the approximate electronic energy in the Thomas-Fermi-Dirac model, J is the electron-electron repulsion represented by a Coulombic term, and K represents an exchange expression.

Despite the improvements, the description of the system as a uniform electron gas model in the Thomas-Fermi and Thomas-Fermi-Dirac models is a reasonable approximation for specific metallic systems but is poor for atoms and molecules. Another relevant point is that both models predict that molecules are unstable. Thus, due to calculation errors and the lack of rigor in their theoretical foundation (*i.e.*, without establishing the variational principle), both models had little impact on chemistry. This scenario was changed with the two theorems established in 1964 by Hohenberg and Kohn [25], which are the basis of the modern density functional theory and allowed it to become a legitimate quantum chemistry methodology.

The first theorem (*i.e.*, the Hohenberg–Kohn existence theorem) demonstrates that the ground-state properties of many-electron systems can be determined by ground state electron

density function [25]. This theorem says that there is a functional that can be used to compute molecular properties from the electron density. It does not say, however, how to find it and what answers of the functional will be at least approximate, according to Eq. (2.49):

$$F\left[\rho_0\right] = E\left[\rho_0\right] = E_0. \tag{2.49}$$

The second theorem (*i.e.*, the Hohenberg–Kohn variational theorem) is analogous to the variational wave function principle for the Hartree-Fock method [25]. The resulting electronic energy will be greater than (or equal to) the true ground state energy for a trial electron density. That is, the second theorem proves that the correct electron density minimizes functional energy, according to Eq.(2.50):

$$F\left[\rho_{\text{trial}}\right] = E\left[\rho_{\text{trial}}\right] \geqslant E_0, \tag{2.50}$$

where ρ_{trial} is a trial electron density (ground-state).

A key step towards determining the unknown functional of the Hohenberg-Kohn theorems was achieved by Kohn and Sham [26] in 1965, who developed a method to calculate the electronic energy of the system from the electron density. For this, Kohn and Sham [26] introduced the Kohn-Sham orbitals, which are auxiliary quantities related to electron density. The Kohn-Sham orbitals follow the precepts of orthonormality of the spin orbitals of the Hartree-Fock method for constructing the Slater determinant. For a system of "N" electronic states, the electron density can be related to the Kohn-Sham orbitals through Eq. (2.51):

$$\rho\left(\mathbf{r}\right) = \sum_{i=1}^{N} \left|\chi_i^{\text{KS}}\right|^2, \tag{2.51}$$

where χ_i^{KS} are Kohn-Sham orbitals.

In the Kohn and Sham [26] formalism, electrons are assumed to be non-interacting and subject to an external potential. Thus, the electronic energy functional (considering the Born-Oppenheimer approximation) can be described in terms of the Kohn-Sham orbitals and the electron density by Eq. (2.52) in Dirac or bra-ket notation for integrals:

$$\begin{aligned} E\left[\rho\left(\mathbf{r}\right)\right] = & \underbrace{\sum_{i=1}^{N}\left(\left\langle\chi_i^{\text{KS}}\right| - \frac{1}{2}\nabla_i^2\left|\chi_i^{\text{KS}}\right\rangle\right)}_{T_{\text{e}}\left[\rho\left(\mathbf{r}\right)\right]} + \underbrace{\sum_{i=1}^{N}\left\langle\chi_i^{\text{KS}}\right|\frac{1}{2}\int\frac{\rho\left(\mathbf{r}'\right)}{\left|\vec{\mathbf{r}}_{\mathbf{i}} - \mathbf{r}'\right|}d\mathbf{r}'\left|\chi_i^{\text{KS}}\right\rangle}_{V_{\text{ee}}\left[\rho\left(\mathbf{r}\right)\right]} \\ & \underbrace{+\left\langle\chi_i^{\text{KS}}\right| - \sum_{a=1}^{M}\frac{Z_a}{\left|\mathbf{r}_i - \mathbf{r}_a\right|}\left|\chi_i^{\text{KS}}\right\rangle\Big)}_{V_{\text{en}}\left[\rho\left(\mathbf{r}\right)\right]} + \underbrace{E_{\text{xc}}\left[\rho\left(\mathbf{r}\right)\right]}_{\substack{\Delta T_{\text{e}}\left[\rho\left(\mathbf{r}\right)\right] \\ +\Delta V_{\text{ee}}\left[\rho\left(\mathbf{r}\right)\right]}}, \end{aligned} \tag{2.52}$$

where T_{e} refers to the kinetic energy of non-interacting electrons functional, V_{ee} refers to electron-electron repulsion energy functional, V_{en} refers to electron-nucleus attraction energy functional, E_{xc} is the exchange-correlation energy functional, ΔT_{e} refers to a correction term in kinetic energy

due to the fact that there is interaction between electrons, and ΔV_{ee} is a term that represents non-classical corrections in the electron-electron repulsion energy.

Regarding Eq. (2.52) previously presented, what determines an accurate calculation of the electronic energy is to have a good approximation for the exchange-correlation energy functional (*i.e.*, E_{xc}). There are several approaches for computing the exchange-correlation energy functional, the two most common being the local density approximation and the generalized gradient approximation.

Similar to the Hartree-Fock method, in density functional theory, the objective is then to determine the Kohn-Sham orbitals to minimize the electronic energy given by Eq. (2.52) through the undetermined Lagrange multipliers method. For this, a Lagrange functional is constructed and given by Eq. (2.53):

$$\mathscr{L}\left[\rho\left(\mathbf{r}\right)\right] = E\left[\rho\left(\mathbf{r}\right)\right] - \sum_{i=1}^{N}\sum_{j\neq i}^{N}\lambda_{ij}\left(\left\langle \chi_i^{\mathrm{KS}}|\chi_j^{\mathrm{KS}}\right\rangle - \delta_{ij}\right). \tag{2.53}$$

For the electronic energy functional to be minimal, it is also necessary to apply an infinitesimal variation in the Kohn-Sham orbitals and set the variation of the Lagrange functional equal to zero, solving Eq. (2.55):

$$\delta\mathscr{L}\left[\rho\left(\mathbf{r}\right)\right] = \delta E\left[\rho\left(\mathbf{r}\right)\right] - \sum_{i=1}^{N}\sum_{j\neq i}^{N}\lambda_{ij}\delta\left\langle \chi_i^{\mathrm{KS}}|\chi_j^{\mathrm{KS}}\right\rangle = 0. \tag{2.54}$$

Through algebraic manipulations in Eq. (2.53) (whose steps can be readily found in the literature), the Kohn-Sham equations are obtained according to Eq. (2.55):

$$\underbrace{\left\{-\frac{1}{2}\nabla_i^2 + \int d\mathbf{r}'\,\frac{\rho\left(\mathbf{r}'\right)}{|\mathbf{r}_i-\mathbf{r}'|} - \sum_{a=1}^{M}\frac{Z_a}{|\mathbf{r}_i-\mathbf{r}_a|} + \frac{\delta E_{\mathrm{xc}}}{\delta\rho\left(\mathbf{r}\right)}\right\}}_{\hat{h}^{\mathrm{KS}}(i)}\chi_i^{\mathrm{KS}}\left(\mathbf{r}\right) = \sum_{j\neq i}^{N}\lambda_{ij}\chi_j^{\mathrm{KS}}\left(\mathbf{r}\right). \qquad \text{for } i = 1,2,\cdots,N \tag{2.55}$$

The set of N-coupled Kohn-Sham equations given by Eq. (2.55), similarly to the Hartree-Fock method, can also be solved by expanding the Kohn-Sham orbitals in terms of basis functions, as described by Eq. (2.56):

$$\chi_i^{\mathrm{KS}} = \sum_{u=1}^{k} c_{u,i}\phi_u, \qquad \text{for } i = 1,2,\cdots,N; \tag{2.56}$$

where k is the number of basis functions, c and ϕ are expansion coefficients to be determined, and subscript i is a label for electronic states.

2.1.5 Intermolecular interactions by ab initio calculations

Intermolecular interactions consist of all interactions between molecules that are significantly weaker than the chemical bond (*i.e.*, all non-covalent interactions). Two main possibilities in calculating energy from intermolecular interactions in quantum chemistry are i) supermolecular and ii) perturbational methods [28–30, 55].

2.1.5.1 Supermolecular methods

Neither the Hartree-Fock method nor density functional theory adequately describes London interactions due to the simplifications considered in both models. The problem is that in the Hartree-Fock method, electron-electron repulsion is treated as an average. In density functional theory, the limitation is related to how the exchange-correlation energy functional is described. Thus, dispersion-corrected density functional theory [31–33], variants of Møller-Plesset perturbation theory [34, 35], and coupled cluster approximation have been used as supermolecular methods [35–37].

In the supermolecular methods, the interaction energy is obtained by the difference between the energy of the dimer and the sum of the individual monomers' energies, according to Eq. (2.57):

$$E_{\text{int}}^{\text{supermol}} = E_{AB} - (E_A + E_B) \tag{2.57}$$

where $E_{\text{int}}^{\text{supermol}}$ is the interaction energy in supermolecular methods, E_{AB} is the energy of dimer composed by monomers A and B, E_A and E_B are the energies of the isolated monomers A and B, respectively.

In practical calculation, the supermolecular methods contain the basis set superposition error resulting from the nonphysical lowering of the monomer's energy in the dimer's calculation due to "borrowing" the basis set from the interacting partner. One way to minimize the effects of basis set superposition is through the counterpoise method. In the counterpoise method, the calculated energy of one monomer considers the energy of the other monomer that constitutes the dimer and vice versa [28–30, 55].

2.1.5.2 Perturbational methods

Symmetry-adapted perturbation theory (SAPT) [28–30, 55] is one of the most widespread perturbational methods. In SAPT methods, the Hamiltonian is also divided into an unperturbed and a perturbation operator, according to the Rayleigh-Schrödinger's perturbation theory. In addition to the perturbation operator, a complementary perturbation for each monomer is added as a correction to the electron correlation description. For a dimer AB, the Hamiltonian

can be obtained from the isolated monomers A and B from Eq. (2.58):

$$\hat{\mathscr{H}}_{AB} = \underbrace{\left(\hat{\mathscr{F}}_A + \varsigma\hat{\mathscr{W}}_A\right)}_{\hat{\mathscr{H}}_A} + \underbrace{\left(\hat{\mathscr{F}}_B + \xi\hat{\mathscr{W}}_B\right)}_{\hat{\mathscr{H}}_B} + \lambda\hat{\mathscr{V}} \tag{2.58}$$

where $\hat{\mathscr{H}}$ is the Hamiltonian, $\hat{\mathscr{F}}$ is the Fock operator of the monomer, $\hat{\mathscr{W}}$ is the Møller-Plesset fluctuation potential of the monomer (it describes the intramolecular electron correlation), $\hat{\mathscr{V}}$ is the interaction potential between the monomers of the dimer, ς and ξ are intramolecular perturbation orders, λ is the intermolecular perturbation order (and define the order of perturbation), subscript AB corresponds to the dimer formed by the A and B monomers, and subscripts A and B represent the isolated monomers.

As well as supermolecular methods such as Møller-Plesset perturbation theory and coupled cluster approximation, SAPT also presents a sophistication in treating electronic correlation. In SAPT, the interaction energy is defined as a perturbation series of polarization and exchange terms given by Eq. (2.59):

$$E_{\text{int}}^{\text{SAPT}} = \sum_{n=1}^{\infty}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\left(E_{\text{pol}}^{nkl} + E_{\text{exch}}^{nkl}\right), \tag{2.59}$$

where $E_{\text{int}}^{\text{SAPT}}$ is the interaction energy in SAPT methods, E_{pol} is a term derived from the polarization expansion, E_{exch} are terms resulting from the antisymmetry of the wave function, n represents the order in $\hat{\mathscr{V}}$ in Eq. (2.58), and k and l represent the orders in $\hat{\mathscr{W}}_A$ and $\hat{\mathscr{W}}_B$ in Eq. (2.58), respectively.

If compared to supermolecular methods, SAPT methods preserve the nature of intermolecular interactions. The key point of SAPT is that within it the interaction energy is decomposed into four fundamental physical contributions verified in all intermolecular interactions: electrostatics, exchange, induction, and dispersion.

Electrostatics corresponds to Coulombic multipole-multipole-type interactions and the interpenetration of charge clouds. The exchange represents the action of a repulsive force and is related to monomer wavefunction overlap and the fermionic antisymmetry requirements of the dimer wavefunction. Induction is related to polarization effects from each monomer's response to the other's electric field and charge transfer. Dispersion represents the action of an attractive force resulting from the dynamic correlation between electrons on one monomer and those on another.

The perturbation series of Eq. (2.59) can be truncated according to the nature of the molecular system studied and the desired accuracy in the calculations. Several levels of SAPT are established in the literature. SAPT terms depend on the four fundamental physical contributions (*i.e.*, electrostatics, exchange, induction, and dispersion). They are labeled according to the order terms of Eq. (2.59) as $E^{(vw)}$, where $v = n$ and $w = k + l$.

The simplest and most computationally efficient SAPT level is SAPT0. In SAPT0, the energy of the monomers is treated by the Hartree-Fock method, and for the treatment of the dimer by Hartree-Fock, dispersion terms from the second-order perturbation theory are added to the electrostatic, exchange, and induction terms. The interaction energy of SAPT0 is defined by Eq. (2.60):

$$\begin{aligned} E_{\text{int}}^{\text{SAPT0}} &= \left[E_{\text{elst}}^{(10)}\right]_{\text{elst}} + \left[E_{\text{exch}}^{(10)}\right]_{\text{exch}} + \left[E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + \delta E_{\text{HF}}^{(2)}\right]_{\text{ind}} \\ &\quad + \left[E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}\right]_{\text{disp}} \\ &= E_{\text{int}}^{\text{HF}} + \left[E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}\right]_{\text{disp}}, \end{aligned} \tag{2.60}$$

[13]

where $E_{\text{int}}^{\text{SAPT0}}$ is the interaction energy in the SAPT0 level, superscript (vw) is related to the order of terms in Eq. (2.59) (*i.e.*, $v = n$ and $w = k + l$), superscript or subscript HF refers to Hartree-Fock method, subscripts elst, exch, ind, and disp are related to the electrostatics, exchange, induction, and dispersion SAPT contributions, respectively; subscript resp indicates the presence of orbital relaxation effects caused by the other monomer, the term $\delta E_{\text{HF}}^{(2)}$ is the Hartree-Fock second-order correction and is defined by by the difference between the interaction energy of the Hartree-Fock method and the SAPT interaction energy without the dispersion contribution (see Eq. (2.60) footnotes for details), and the square brackets [] are used to separate SAPT terms into the four fundamental physical contributions.

Since, as shown in Eq. (2.60), SAPT0 neglects intramolecular electron correlation terms (*i.e.*, $w = 0$ in the terms $E^{(vw)}$), its accuracy is limited. The correction in some of the SAPT0 terms led to a new method called sSAPT0, which improved interaction energy calculation results for some instances, such as strongly interacting systems. In sSAPT0, the interaction energy is defined by Eq. (2.61):

$$\begin{aligned} E_{\text{int}}^{\text{sSAPT0}} &= \left[E_{\text{elst}}^{(10)}\right]_{\text{elst}} + \left[E_{\text{exch}}^{(10)}\right]_{\text{exch}} + \left[E_{\text{ind,resp}}^{(20)} + p_{\text{EX}}(3.0)\, E_{\text{exch-ind,resp}}^{(20)} + \delta E_{\text{HF}}^{(2)}\right]_{\text{ind}} \\ &\quad + \left[E_{\text{disp}}^{(20)} + p_{\text{EX}}(3.0)\, E_{\text{exch-disp}}^{(20)}\right]_{\text{disp}}, \end{aligned} \tag{2.61}$$

[14]

where $E_{\text{int}}^{\text{sSAPT0}}$ is the interaction energy in the sSAPT0 level, $p_{\text{EX}}(\alpha)$ is a correction term[15], and the term $\delta E_{\text{HF}}^{(2)}$ are Hartree-Fock second-order correction (see Eq. (2.61) footnotes for details).

With the insertion of intramolecular electron correlation terms (*i.e.*, $w > 0$ in the terms $E^{(vw)}$), other SAPT levels more sophisticated than SAPT0 and sSAPT0 emerge. Among these levels one finds SAPT2, SAPT2+, SAPT2+(3), and SAPT2+3, whose interaction energies can be computed according to the set of Eqs (2.62)-(2.65):

---

[13] In Eq. (2.60): $\delta E_{\text{HF}}^{(2)} = E_{\text{int}}^{\text{HF}} - \left(\left[E_{\text{elst}}^{(10)}\right]_{\text{elst}} + \left[E_{\text{exch}}^{(10)}\right]_{\text{exch}} + \left[E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)}\right]_{\text{ind}}\right)$

[14] In Eq. (2.61): $\delta E_{\text{HF}}^{(2)} = E_{\text{int}}^{\text{HF}} - \left(\left[E_{\text{elst}}^{(10)}\right]_{\text{elst}} + \left[E_{\text{exch}}^{(10)}\right]_{\text{exch}} + \left[E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)}\right]_{\text{ind}}\right)$.

[15] $p_{\text{EX}}(\alpha) = \left(\dfrac{E_{\text{exch}^{10}}}{E_{\text{exch}^{10}}(S^2)}\right)^{\alpha}$, with the factor $p_{\text{EX}}(\alpha)$ scaled to $\alpha = 1.0$ in SAPT0 and to $\alpha = 3.0$ in sSAPT0 and (S^2) indicates that the energy term is evaluated using the S^2 approximation.

$$E_{\text{int}}^{\text{SAPT2}} = E_{\text{int}}^{\text{SAPT0}} + \left[E_{\text{elst,resp}}^{(12)}\right]_{\text{elst}} + \left[E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)}\right]_{\text{exch}} + \left[{}^{\text{t}}E_{\text{ind}}^{(22)} + {}^{\text{t}}E_{\text{exch-ind}}^{(22)}\right]_{\text{ind}}, \tag{2.62}$$

$$E_{\text{int}}^{\text{SAPT2+}} = E_{\text{int}}^{\text{SAPT2}} + \left[E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)}\right]_{\text{disp}}, \tag{2.63}$$

$$E_{\text{int}}^{\text{SAPT2+(3)}} = E_{\text{int}}^{\text{SAPT2+}} + \left[E_{\text{elst,resp}}^{(13)}\right]_{\text{elst}} + \left[E_{\text{disp}}^{(30)}\right]_{\text{disp}}, \tag{2.64}$$

$$\begin{aligned} E_{\text{int}}^{\text{SAPT2+3}} = E_{\text{int}}^{\text{SAPT2+(3)}} &+ \left[E_{\text{ind,resp}}^{(30)} + E_{\text{exch-ind}}^{(30)} - \delta E_{\text{HF}}^{(2)} + \delta E_{\text{HF}}^{(3)}\right]_{\text{ind}} \\ &+ \left[E_{\text{exch-disp}}^{(30)} + E_{\text{ind-disp}}^{(30)} + E_{\text{exch-ind-disp}}^{(30)}\right]_{\text{disp}}, \end{aligned} \tag{2.65}$$

[16]

where $E_{\text{int}}^{\text{SAPT2}}$, $E_{\text{int}}^{\text{SAPT2+}}$, $E_{\text{int}}^{\text{SAPT2+(3)}}$, and $E_{\text{int}}^{\text{SAPT2+3}}$ are the interaction energies in the SAPT2, SAPT2+, SAPT2+(3), and SAPT2+3 levels, respectively; $\delta E_{\text{HF}}^{(2)}$ and $\delta E_{\text{HF}}^{(3)}$ are Hartree-Fock second and third-order corrections (see Eq. (2.65) footnotes for further details), and superscript *t* indicates true correlation contributions.

To the levels of SAPT beyond SAPT2 previously described, an additional correction can be employed for certain types of systems (*e.g.*, electrostatically dominated complexes). The correction is based on the Møller-Plesset supermolecular interaction energy, originating the methods SAPT2+δMP2, SAPT2+(3)δMP2, and SAPT2+3δMP2, which interaction energies can be computed according to the set of Eqs (2.66)-(2.68):[17]

$$E_{\text{int}}^{\text{SAPT2+}\delta\text{MP2}} = E_{\text{int}}^{\text{SAPT2+}} + \left[\delta E_{\text{MP2}}\right]_{\text{ind}}, \tag{2.66}$$

$$E_{\text{int}}^{\text{SAPT2+(3)}\delta\text{MP2}} = E_{\text{int}}^{\text{SAPT2+(3)}} + \left[\delta E_{\text{MP2}}\right]_{\text{ind}}, \tag{2.67}$$

$$E_{\text{int}}^{\text{SAPT2+3}\delta\text{MP2}} = E_{\text{int}}^{\text{SAPT2+3}} + \left[\delta E_{\text{MP2}}\right]_{\text{ind}}, \tag{2.68}$$

wherein $E_{\text{int}}^{\text{SAPT2+}\delta\text{MP2}}$, $E_{\text{int}}^{\text{SAPT2+}\delta\text{MP2}}$, $E_{\text{int}}^{\text{SAPT2+(3)}\delta\text{MP2}}$, $E_{\text{int}}^{\text{SAPT2+(3)}\delta\text{MP2}}$, $E_{\text{int}}^{\text{SAPT2+3}\delta\text{MP2}}$, and $E_{\text{int}}^{\text{SAPT2+3}\delta\text{MP2}}$ are the interaction energies in the SAPT2+δMP2, SAPT2+MP2, SAPT2+(3)δMP2, SAPT2+(3)MP2, SAPT2+3δMP2, and SAPT2+3MP2 levels, respectively; and δE_{MP2} is the second-order Møller-Plesset interaction energy correction (see Eqs. (2.66)-(2.68) footnotes for details).

2.1.6 Basis set for intermolecular interactions

The numerical solution of the Schrödinger's equation through computational codes by methods based on computing the electronic wave function or electron density needs to use the

[16] In Eq. (2.65):
$\delta E_{\text{HF}}^{(2)} = E_{\text{int}}^{\text{HF}} - \left(\left[E_{\text{elst}}^{(10)}\right]_{\text{elst}} + \left[E_{\text{exch}}^{(10)}\right]_{\text{exch}} + \left[E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)}\right]_{\text{ind}}\right)$ and
$\delta E_{\text{HF}}^{(3)} = E_{\text{int}}^{\text{HF}} - \left(\left[E_{\text{elst}}^{(10)}\right]_{\text{elst}} + \left[E_{\text{exch}}^{(10)}\right]_{\text{exch}} + \left[E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + E_{\text{ind,resp}}^{(30)} + E_{\text{exch-ind,resp}}^{(30)}\right]_{\text{ind}}\right)$

[17] In Eqs. (2.66)-(2.68): $\delta E_{\text{MP2}} \equiv E_{\text{int}}^{\text{MP2}} - E_{\text{int}}^{\text{SAPT2}}$, wherein $E_{\text{int}}^{\text{MP2}}$ is the MP2 interaction energy (see Eq. (2.39) for further details) counterpoise corrected and $E_{\text{int}}^{\text{SAPT2}}$ is the interaction energy in the SAPT2 level (see Eq. (2.62) for further details)

basis set approximation [56,57]. A basis set is a set of known mathematical functions (also called basis functions) used to represent atomic orbitals. Physically, basis functions can describe the electron distribution around a neutral atom and, when combined linearly, describe the electron distribution in a molecule. Two of the basis functions widely used are Slater and Gaussian type atomic orbitals[18]. Both approaches are expressed as a function of separate variables composed of a radial and an angular part. The mathematical description of the Slater and Gaussian type atomic orbitals are given in spherical coordinates by Eqs. (2.69) and (2.70), respectively:

$$\phi_{\mathrm{STO}}(r,\theta,\varphi) = N\underbrace{r^{n-1}e^{-\zeta r}}_{\text{radial part}}\underbrace{Y_{l,m}(\theta,\varphi)}_{\text{angular part}}, \tag{2.69}$$

$$\phi_{\mathrm{GTO}}(r,\theta,\varphi) = N\underbrace{r^{2n-2-l}e^{-\zeta r^2}}_{\text{radial part}}\underbrace{Y_{l,m}(\theta,\varphi)}_{\text{angular part}}, \tag{2.70}$$

where ϕ is the basis function, variables r, θ, and φ are spherical coordinates (where r represents the distance between electron and atomic nucleus); N is a normalization constant, ζ is the orbital exponent, $Y_{l,m}$ are spherical harmonic functions, variables n, l, and m are related to the quantum numbers (principal, azimuthal, and magnetic, respectively), and subscripts STO and GTO are related to Slater and Gaussian type atomic orbitals.

From Eqs. (2.69) and (2.70), the difference between Slater and Gaussian-type atomic orbitals lies in the representation of the radial part. Gaussian-type atomic orbitals show a quadratic dependence on the distance between the electron and atomic nucleus (*i.e.*, it exhibits a quadratic dependence on the variable r). This feature of Gaussian-type atomic orbitals implies shortcomings such as a poor mathematical description of the electron distribution around the nucleus (*i.e.*, it has a slope at $r \to 0$) and a very rapid decay of the electron distribution with increasing distance from the nucleus (*i.e.*, it drops very quickly in the direction of $r \to \infty$). Gaussian-type atomic orbitals have the advantage of having analytical solutions, and by combining several Gaussian-type atomic orbitals, the shortcomings can be minimized. These characteristics make Gaussian-type atomic orbitals more commonly used compared to Slater-type atomic orbitals [56, 57].

Selecting a basis set is an essential step in *ab initio* calculations. To obtain the exact energy of a given system, *ab initio* calculations must be performed at the complete basis set (CBS) limit. Performing *ab initio* complete basis set calculations implies using an infinite number of basis functions, which is impossible. Obtaining good chemical accuracy with the use of a small basis set, however, is possible. The number of basis functions in a basis set depends on factors such as the number of core and valence atomic orbitals and whether the atom is light or heavy [56, 57].

A minimal or single zeta (SZ) basis set describes each atomic orbital by only one basis function. A basis set in which two basis functions are used to represent each atomic orbital

[18] Inspired by the form of the bound state hydrogen wave functions, the initially used radial basis functions were exponential functions of the form of Slater-Type Orbitals (STOs)

is called double-zeta (DZ). When using three basis functions to represent each atomic orbital, the basis set is called triple-zeta (TZ). For four basis functions representing each atomic orbital, the basis set is called quadruple-zeta (QZ), and so on. Chemical accuracy can be further improved significantly with extended basis sets. There are several different types of extended basis sets; among them are: split valence, polarized, diffuse, and correlation-consistent basis sets [56, 57].

Split valence basis sets use more than one basis function to represent the valence atomic orbitals, since in molecules the valence electrons are more important because they are responsible for forming the chemical bonds. Split valence basis sets that use two, three, or four basis functions are called valence double (VDZ), triple (VTZ), or quadruple-zeta (VQZ) basis sets, respectively. Polarized basis sets include polarization functions to describe the distortions of the electronic cloud in a molecular environment, which are essential in the description of chemical bonds and correspond to additional functions with an angular moment different from that presented by the original basis set. Diffuse basis sets include diffuse functions to describe the behavior of systems in which electrons are weakly bound (*e.g.*, anions or excited states), since the electron cloud in these systems tends to expand. Correlation consistent basis sets contain correlating functions that reduce the correlation energy in post Hartree-Fock methods (*e.g.*, Møller-Plesset and coupled cluster methods) and were developed to smoothly converge to the complete basis set limit [56, 57].

One of the most used correlation consistent basis sets was developed by Dunning [58] in 1989. Dunning basis sets are commonly expressed in the form "aug-cc-pV*n*Z", where the prefix "aug" denotes the presence of diffuse functions and is optional, "cc" indicates that it is correlation consistent, "p" indicates the presence of polarization functions and "V*n*Z " indicates that it has "*n*" zeta basis functions in valence atomic orbitals (*e.g.*, *n* = D, T, and Q for double, triple, and quadruple-zeta, respectively). Intermolecular interactions (*i.e.*, non-covalent interactions) show a very subtle variation in energy and electronic redistribution [28–30, 55]. Dunning basis sets are an ideal choice to describe intermolecular interactions as they have both the necessary polarization and diffuse functions [28–30, 55].

2.1.7 Molecular dynamics

Within the molecular scale, the description of molecules comprises studying the dynamics of many-body systems. Among the computational methods available for this purpose, molecular dynamics is one of the most commonly used. Molecular dynamics is a versatile computational technique grounded on statistical mechanics, in which molecules are treated as a set of particles subject to an interaction potential [12–15, 27].

The main pillar of molecular dynamics simulations is to know the interaction potential between particles. This interaction potential between particles is called force field. The resulting force acting on the particles can be associated with the interaction potential that is

dependent on the position of the particles, according to Eq. (2.71):

$$\mathbf{F}_i(t) = -\nabla U\left(\mathbf{r}_1(t), \mathbf{r}_2(t), \cdots, \mathbf{r}_N(t)\right), \tag{2.71}$$

where $\mathbf{F}_i$ is the force vector acting on the particle i, t is the time, ∇ is the gradient operator $\left(i.e., \nabla = \frac{\partial}{\partial x}\hat{\mathbf{i}} + \frac{\partial}{\partial y}\hat{\mathbf{j}} + \frac{\partial}{\partial z}\hat{\mathbf{k}}\right)$, U is the potential energy, $\mathbf{r}_i$ is the position vector for particles, and subscripts $1, 2, \cdots, N$ are labels for particles.

To understand the behavior of particles at the molecular level, it is necessary to compute how the particles positions evolve over time. For this, classical mechanics is used, in particular, Newton's equations of motion. Using Newton's equations of motion, if the resulting force acting on the particle is known, it is possible to calculate the change in momentum of the particle over time and, consequently, the change in its position over time, according to Eq. (2.72):

$$\mathbf{F}_i(t) = \frac{\mathrm{d}\mathbf{p}_i(t)}{\mathrm{d}t} = m_i \frac{\mathrm{d}^2\mathbf{r}_i(t)}{\mathrm{d}t^2}, \tag{2.72}$$

where $\mathbf{p}_i$ is the momentum of particle i and m_i is the mass of the particle i.

2.1.7.1 Force fields and parameterization

In molecular dynamics, a force field is an empirical potential energy function capable of describing interactions between particles at the molecular level. Based on molecular mechanics, the first force fields were developed in the 1960s to create models able to predict properties of isolated small organic molecules (*e.g.*, molecular structure, vibrational spectra, and enthalpies) [27]. Many of these initial force fields continue to be used and developed to this day as MM potentials (*e.g.*, MM2, MM3, MM4) [59–66]. Literature also accounts for many other quite popular force fields such as DREIDING [67], UFF [68, 69], CHARMM [70], AMBER [71], GROMOS [72], OPLS [73], and COMPASS [74]. Each of these previously mentioned force fields differs in their applications and accuracy to represent the systems for which they were developed.

The force field (or total potential energy surface) is a set of equations dependent only on the particle position coordinates in a given system. The total potential energy can be described as a sum of the energy terms of the bonded and non-bonded interactions, according to Eq. (2.73):

$$U_{\text{total}} = U_{\text{bonded}} + U_{\text{non-bonded}}, \tag{2.73}$$

where U_{total} is the total potential energy, U_{bonded} is the bonded potential energy, and $U_{\text{non-bonded}}$ is the non-bonded potential energy.

In the force fields most commonly used in the literature [12–15, 27], the bonded potential represented in Eq. (2.74) typically consists of four main contributions which are i) bond stretching (*i.e.*, due to the stretching of two adjacent bonded particles), ii) angular bending

(*i.e.,* due to angular bending of three adjacent bonded particles), iii) dihedral torsion (*i.e.*, due to dihedral torsion of four adjacent bonded particles), and iv) improper torsion or out-of-plane deformation (also verified in four adjacent bonded particles):

$$U_{\text{bonded}} = \underbrace{\sum_{\text{bonds}} U_{\text{bond}}\left(b_{ij}\right)}_{\text{bond stretching}} + \underbrace{\sum_{\text{angles}} U_{\text{angle}}\left(\theta_{ijk}\right)}_{\text{angle bending}} + \underbrace{\sum_{\text{dihedrals}} U_{\text{dihedral}}\left(\varphi_{ijkl}\right)}_{\text{dihedral torsion}} + \underbrace{\sum_{\text{impropers}} U_{\text{improper}}\left(\omega_{ijkl}\right)}_{\text{improper torsion}}, \tag{2.74}$$

where U_{bond} is the bond stretching potential energy, b_{ij} is the distance between the bonded particles i and j, U_{angle} is the angle bending potential energy, θ_{ijk} is the angle formed by the bonded particles i, j, and k; U_{dihedral} is the dihedral torsion potential energy, φ_{ijkl} is the dihedral angle formed by the bonded particles i, j, k, and l; U_{improper} is the improper torsion potential energy, and ω_{ijkl} is the angle formed between the particle i and the plane formed by the particles j, k and l with all these particles bonded.

Harmonic potentials are appropriate to represent interactions related to bond stretching, angular bending, and improper torsion because the deviations from the bonds and angles between atoms in these cases are less sensitive when compared to the lower energy values. On the other hand, the dihedral torsion is often modeled by trigonometric functions to represent the periodicity of the potential associated with interactions between neighboring atoms due to rotations around the central bond. The simplest ways found in the literature [12–15, 27] to represent bond stretching, angle bending, dihedral torsion, and improper torsion potential energies are given by Eqs. (2.75), (2.76), (2.77), and (2.78), respectively:

$$U_{\text{bond}}\left(b_{ij}\right) = k_b\left(b_{ij} - b_{ij}^0\right)^2, \tag{2.75}$$

$$U_{\text{angle}}\left(\theta_{ijk}\right) = k_\theta\left(\theta_{ijk} - \theta_{ijk}^0\right)^2, \tag{2.76}$$

$$U_{\text{dihedral}}\left(\varphi_{ijk}\right) = \frac{1}{2}k_\varphi\left[1 + \cos\left(n\varphi_{ijkl} - \xi\right)\right], \tag{2.77}$$

$$U_{\text{improper}}\left(\omega_{ijkl}\right) = k_\omega\omega_{ijkl}^2, \tag{2.78}$$

where k_b is the bond stretching force constant, b_{ij}^0 is the equilibrium bond distance of particles i and j, k_θ is the angle bending force constant, θ_{ijk}^0 is the equilibrium angle formed by the particles i, j, and k; k_φ is the dihedral torsion force constant, n is the multiplicity (*i.e.*, is the number points of minimum energy over a complete rotation of the bond formed by the particles j and k), ξ is the phase angle, U_{improper} is the improper torsion potential energy, k_ω is the improper torsion force constant.

For the non-bonded potential, one of the models most commonly found in the force fields available in the literature [12–15, 27] consist of i) van der Waals and ii) electrostatic

contributions, as described by (2.79):

$$U_{\text{non-bonded}} = \sum_{\text{pairs}} U_{\text{vdw}}\left(r_{ij}\right) + \sum_{\text{pairs}} U_{\text{elst}}\left(r_{ij}\right), \tag{2.79}$$

where U_{vdw} is the van der Waals potential energy, U_{elst} is the electrostatic potential energy, r_{ij} is the distance of non-bonded particles i and j.

The contribution of van der Waals interactions arises from two contributions: repulsion and attraction between two nuclei of adjacent atoms. The repulsive term becomes predominant at short distances due to repulsion caused by overlapping electron clouds, while the attractive term is related to London interaction, which is essentially a dipole-dipole interaction. Electrostatic interaction contributions describe specific electrical interactions as expressed by Coulomb's law. The Lennard-Jones potential, which is one of the most commonly used to represent van der Waals interactions, and the Coulomb potential, used to represent electrostatic interactions, are given by Eqs. (2.80) and (2.81), respectively:

$$U_{\text{vdw}}\left(r_{ij}\right) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{2.80}$$

$$U_{\text{elst}}\left(r_{ij}\right) = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}}, \tag{2.81}$$

where ε_{ij} is the depth of the potential energy well of non-bonded particles i and j, σ_{ij} is the distance at which the potential is zero for the pairs of non-bonded particles i and j, r_{ij} is the distance between the centers of non-bonded particles i and j, ε_0 is the vacuum permittivity, and variables q_i and q_j are the electrical charges of non-bonded particles i and j.

For mixtures properties, the non-bonded potential parameters for pure components must be used to compute mixture-specific parameters through combining rules [12–15, 27]. Among the existing combining rules, one of the best known is the Lorentz-Berthelot. Lorentz-Berthelot combining rules are given by Eqs. (2.82) and (2.83), respectively:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}, \tag{2.82}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}, \tag{2.83}$$

where σ is the distance at which the potential are zero for a pair of non-bonded particles, ε is the depth of the potential energy well for a pair of non-bonded particles, and subscripts i and j are relative to non-bonded particles.

The bonded and non-bonded potential energy contributions previously presented are present in almost all force fields in the literature [12–15, 27]. Other additional terms may be found to better represent specific chemical systems. In the bonded potential, cross-terms can be derived. These terms usually describe the coupling between bond stretching, angular bending, dihedral torsion, and improper torsion. They bring corrections to the bonded potential energy

and allow it to better reproduce intramolecular vibrations spectra. An example of a potential with cross-terms is given by Eq. (2.84):

$$U_{\text{cross}}\left(b_{ij}, b_{jk}, \theta_{ijk}, \cdots\right) = \sum_{\text{bonds}} U_{\text{bond-bond}}\left(b_{ij}, b_{jk}\right) + \sum_{\text{bonds,angles}} U_{\text{bond-angle}}\left(b_{ij}, b_{jk}, \theta_{ijk}\right) + \cdots, \quad (2.84)$$

[19]

where U_{cross} is a potential energy due to cross-terms, variables b_{ij} and b_{jk} are the bond distance of particles i and j and j and k, respectively; θ_{ijk} is the angle formed by the particles i, j, and k, $U_{\text{bond-bond}}$ is the potential energy due to stretching of two verified bonds between three particles, $U_{\text{bond-angle}}$ is the potential due to stretching of two verified bonds between three particles and the angular bending between these three particles.

Another important energy term addition is the portion related to the polarization effects, which is a non-pairwise additive interaction [12–15, 27]. In specific chemical systems, the inclusion of a polarization potential described by Eq. (2.85) is justified because local electric field effects can lead to the appearance of dipoles:

$$U_{\text{pol}} = -\frac{1}{2}\sum_{i} \mu_i \mathbf{E}_i, \quad (2.85)$$

where U_{pol} is the polarization potential energy, μ_i is an induced point dipole on particle i, $\mathbf{E}_i$ is the electrostatic field on atom i due to charges and induced dipoles.

Atomistic force fields, in which each atom is treated as a particle, are reasonably accurate models for describing molecules. Atomistic force fields, however, may not be suitable for simulating molecules with a large number of atoms (*e.g.*, polymers and proteins) or long-time phenomena (*e.g.*, self-assembly and slow diffusion processes). Coarse-graining methods were created and developed to overcome these limitations. Such methods involve representing a group of atoms in beads. With this kind of simplification, the chemical details of the molecule cannot be fully described. Effective interactions (*i.e.*, bonded and non-bonded interactions) between the beads can be determined to mimic some of the key characteristics found in the original molecule. Coarse-grained force fields allow larger time and size scales in molecular simulations with significant savings in computational resources [12–15, 27].

The process of obtaining force field parameters is called force field parameterization. Two methodologies for obtaining force field parameters are available in the literature, they are: i) top-down and ii) bottom-up. In the top-down methodology, the force field parameters are fitted to reproduce experimental macroscopic properties. In the bottom-up approach, force field parameters are fitted to reproduce results of *ab initio* calculations. While force field parameters

[19] In Eq. (2.84), examples of terms $U_{\text{bond-bond}}$ and $U_{\text{bond-angle}}$ are:
$U_{\text{bond-bond}}\left(b_{ij}, b_{jk}\right) = k_{bb}\left(b_{ij} - b_{ij}^0\right)\left(b_{jk} - b_{jk}^0\right)$, and
$U_{\text{bond-angle}}\left(b_{ij}, b_{jk}, \theta_{ijk}\right) = k_{ba1}\left(b_{ij} - b_{ij}^0\right)\left(\theta_{ijk} - \theta_{ijk}^0\right) + k_{ba2}\left(b_{ij} - b_{ij}^0\right)\left(\theta_{ijk} - \theta_{ijk}^0\right)$
where k_{bb}, k_{ba1}, and k_{ba2} are force constants and superscript 0 is related to equilibrium values.

obtained by the top-down approach have greater representability of experimental data, force field parameters obtained by the bottom-up approach generally have greater transferability and can be applied to systems similar to the parameterized one [12–15, 27].

The literature [12–15, 27] more clearly establishes the methodology for obtaining bonded potential parameters. According to the bottom-up methodology, bonded potential parameters are usually fitted to represent energy variations resulting from *ab initio* calculations. These energies come from the difference between the energy of a molecule with distorted spatial configuration and the energy of the same molecule in equilibrium (*i.e.*, with the spatial structure optimized for minimum energy). A molecule with distorted spatial configuration has undergone variations in the length of chemical bonds and angles between atoms about its chemical structure in equilibrium.

One of the main challenges in developing force fields is obtaining reliable parameters for the non-bonded potential. The direct and simplest route to obtain non-bonded potential parameters for a given system is through the top-down methodology, even though there is a loss in parameter transferability. The methodology, however, for obtaining parameters of the non-bonded potential through *ab initio* calculations by the bottom-up methodology, although it seems well established, lack clear details when described in the literature [12–15, 27]. Parameters of the van der Waals potential are usually adjusted to reproduce the interaction energy between molecules, and this energy can be obtained by supermolecular or perturbational methods [28–30, 55] (for recapitulation, see sections 2.1.5.1 and 2.1.5.2, respectively). Moreover, parameters of the Coulomb potential can also be obtained by *ab initio* calculations by the schemes of Mulliken, ESP, Chelp, and ChelpG, among others [12–15, 27].

2.1.7.2 Molecular dynamics simulations

Molecular dynamics simulations consist of obtaining the trajectory of particles over time by integrating Newton's equations of motion. Obtaining the trajectory of the particles is nothing less than mapping the positions and velocities of the particles over time, and use this information to compute properties at the molecular level of a given system. Some examples of numerical algorithms used to integrate Newton's equations of motion are Verlet's [75, 76], velocity-Verlet [77], and leapfrog [78]. The important point is that the equations of motion must be time-reversible, and respect energy conservation.

Verlet's algorithm was applied in molecular dynamics simulations in the 1960s by Verlet [75, 76]. The basic idea of Verlet's algorithm is to derive the position of particles over time from Taylor expansions, given by Eqs. (2.86) and (2.87)[75, 76]:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\,\delta t + \frac{1}{2}\mathbf{a}(t)\,\delta t^2, \tag{2.86}$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \mathbf{v}(t)\,\delta t + \frac{1}{2}\mathbf{a}(t)\,\delta t^2, \tag{2.87}$$

where the position of the particles over time can be obtained by adding Eqs. (2.86) and (2.87) through Eq. (2.88) and the velocity of the particles can be computed from Eq. (2.89), respectively:

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \mathbf{a}(t)\,\delta t^2, \tag{2.88}$$

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t}, \tag{2.89}$$

where **r**, **v**, and **a** are the position, velocity, and acceleration vectors of the particles, and t is the time.

Although Verlet's algorithm is simple and easy to implement, the velocity of the particles over time is not directly calculated, which can lead to energy conservation problems. Two alternatives to overcome this type of shortcoming are the velocity-Verlet [77] and leapfrog algorithms [78]. In the velocity-Verlet algorithm, the position and velocity of the particles over time are given by Eqs. (2.90) and (2.91), respectively:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\,\delta t + \frac{1}{2}\mathbf{a}(t)\,\delta t^2, \tag{2.90}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \frac{1}{2}\left[\mathbf{a}(t) + \mathbf{a}(t+\delta t)\right], \tag{2.91}$$

and in the leapfrog algorithm, the position and velocity of the particles over time are given by Eqs. (2.92) and (2.93), respectively:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{1}{2}\delta t\right)\delta t, \tag{2.92}$$

$$\mathbf{v}\left(t + \frac{1}{2}\delta t\right) = \mathbf{v}\left(t - \frac{1}{2}\delta t\right) + \mathbf{a}(t). \tag{2.93}$$

In molecular dynamics simulations, the state representation of the studied system (*i.e.*, the set of configurations and properties kept constant during a simulation) is defined by the related statistical ensemble, which is a set of replicas of a thermodynamic system. By the ergodicity hypothesis, which establishes that, in a sufficiently long time, the configurations of a system can be sampled by replicas (*i.e.*, microstates with an energy proportional to the volume of the system initially studied), the ensemble average properties can be replaced by the time averages on the time evolution of one of the replicas [12–15, 27].

Computed properties within molecular dynamics simulations depend on the kind of selected ensemble. Usually, simulation properties such as temperature (T), pressure (P), number of particles (N), energy (E), volume (V), or chemical potential (μ) can be kept constant. The most commonly simulated ensembles are microcanonical or NVE (*i.e.*, where N, V, and E are kept constant), canonical or NVT (*i.e.*, where N, V, and T are kept constant), and isothermal-isobaric or NPT (*i.e.*, where N, P, and T are kept constant), the latter being widely used to simulate experimental conditions (*i.e.*, commonly at constant P and T) [12–15, 27].

2.1.8 Molecular-based, equations of state

Equations of state that consider the molecular nature of the system have reliable predictive and extrapolative power. This class of state equations is called molecular-based, equations of state. Molecular-based equations of state are developed from molecular modeling techniques, an approach where an interaction potential between the interacting molecules of the studied system is considered [1].

In molecular-based, equations of state, the great advantage is that the macroscopic properties are described through molecular contributions, which are considered when building the theoretical foundation of the model. These molecular contributions are constructed with parameters that have physical meaning (*e.g.*, molecular shape information). Within this group of equations of state, one of the most popular is the molecular-based, statistical associating fluid theory (SAFT) equations of state [1, 44, 45].

2.1.8.1 Statistical associating fluid theory (SAFT)

Statistical associating fluid theory (SAFT) is built on fundamental statistical mechanics and is based on Wertheim's first-order thermodynamic perturbation theory (TPT1) [39–43]. In the general formulation of SAFT theory, molecules are modeled as spherical segments (also referred to monomers). Each spherical segment can correspond to an atom, united atom group (*e.g.*, functional group, repetition units in a polymer), or a molecule (*e.g.*, argon, methane). These spherical segments have a repulsive core and multiple attractive sites and can associate to form chains and ring structures [1, 44, 45].

As the spherical segments can be modeled in different ways, the literature has several equations of state based on SAFT theory [1,44,45]. In the SAFT framework, the equation of state of a fluid is a perturbation expansion in which the total Helmholtz free energy corresponds to the sum of different contributions (ideal, segment, chain, and association) at the same temperature and density, according to Eq. (2.94):

$$\frac{A^{\text{tot}}}{Nk_{\text{B}}T} = \frac{A^{\text{ideal}}}{Nk_{\text{B}}T} + \underbrace{\frac{A^{\text{mono}}}{Nk_{\text{B}}T} + \frac{A^{\text{chain}}}{Nk_{\text{B}}T} + \frac{A^{\text{assoc}}}{Nk_{\text{B}}T}}_{\frac{A^{\text{res}}}{Nk_{\text{B}}T}}, \tag{2.94}$$

where A is the Helmholtz free energy, N is the total number of segment units, k_{B} is the Boltzmann constant, T is the temperature, and superscripts ideal, seg, chain, assoc, and res are related to ideal, segment, chain, association, and residual contributions.

The ideal contribution refers to the ideal gas contribution. The segment contribution describes the spherical non-directional interaction of the individual monomers due to repulsion and dispersion (attraction). The chain contribution uses the association of spherical particles described in TPT1 to describe the formation of chains from the segments, where the application

of ring structures involves modifications in the theory. The association contribution describes the formation of complexes through association sites that are a noncentral potential close to the perimeter of the molecule [1, 44, 45].

The literature has extensive reviews [1, 44, 45] that discuss the development of SAFT equations of state, highlighting the differences between the available versions. Variants of SAFT equations of state have been developed and used to investigate thermodynamics from simple hydrocarbons to complex systems (*e.g.*, electrolytes, polymers, pharmaceuticals, confined fluids). Among the existing variants of the SAFT equations of state, some of the principal ones are SAFT-0 [2, 3], SAFT-HR [79, 80], SAFT-VR [81], PC-SAFT [82–84], SAFT-VR Mie [85–87], and SAFT-γ Mie [88]. A brief description of the main differences between these variants is as follows.

In the first version of SAFT (SAFT-0) developed by Chapman *et al.* [2, 3], the Lennard-Jones potential was used to model interactions between segments, given by Eq. (2.95):

$$U^{\mathrm{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{2.95}$$

where U^{LJ} is the Lennard-Jones potential energy, r is the distance between the centers of two segments, ε is the depth of the potential energy well, and σ is the distance at which the potential is zero (independent of temperature).

SAFT by Huang and Radosz [79, 80] (SAFT-HR) was the next model to be developed. In SAFT-HR, the square-well potential of a constant well-width was considered to represent the interactions between segments, according to Eq. (2.96):

$$U^{\mathrm{SW}}(r) = \begin{cases} \infty, & \text{if } r < \sigma \\ -\varepsilon, & \text{if } \sigma \leqslant r < \lambda\sigma \\ 0, & \text{if } r \geqslant \lambda\sigma \end{cases}, \tag{2.96}$$

where U^{SW} is the square-well potential energy, r is the distance between the centers of two segments, σ is the diameter of the segment (independent of temperature), ε is the depth of the potential energy well, and λ is the width of the potential well.

The SAFT with variable range (SAFT-VR) was developed by Gil-Villegas *et al.* [81], and appeared next to allow the treatment of nonconformal molecules. The square-well, Yukawa, and Sutherland potentials were considered to describe the potential for interaction between segments, with the square-well potential being most used [45]. In SAFT-VR, the well-width of the square-well potential of Eq. (2.96) is treated as an adjustable parameter, which means that different types of interactions can be described in different ranges for different molecules (*i.e.*, nonconformal behavior).

The subsequent SAFT model was the Perturbed-Chain SAFT (PC-SAFT), developed by Gross and Sadowski [82–84]. Unlike the previous variants, PC-SAFT applied the dispersion

interactions to the segment chains instead of applying them directly to the segments. In PC-SAFT, the interaction potential between segments was the modified square-well potential described by Chen and Kreglewski [89], according to Eq. (2.97):

$$U^{\mathrm{MSW}}(r) = \begin{cases} \infty, & \text{if } r < (\sigma - s_1) \\ 3\varepsilon, & \text{if } (\sigma - s_1) \leqslant r < \sigma \\ -\varepsilon, & \text{if } \sigma \leqslant r < \lambda\sigma \\ 0, & \text{if } r \geqslant \lambda\sigma \end{cases}, \tag{2.97}$$

where U^{MSW} is the modified square-well potential energy, r is the distance between the centers of two segments, σ is the diameter of the segment (temperature-independent), $s_1 = 0.12\sigma$, ε is the depth of the potential energy well, and λ is the width of the potential well.

SAFT-VR was later extended by Lafitte *et al.* [85–87] to consider the Mie potential, which is a generalization of Lennard-Jones potential, originating the SAFT-VR Mie. In the SAFT-VR Mie equation of state, one of the last incarnations of the SAFT theory, the Mie potential given by Eqs. (2.98) and (2.99) is used to describe the interactions between the segments:

$$U^{\mathrm{Mie}}(r) = C\varepsilon\left[\left(\frac{\sigma}{r}\right)^{\lambda_{\mathrm{rep}}} - \left(\frac{\sigma}{r}\right)^{\lambda_{\mathrm{att}}}\right], \tag{2.98}$$

$$C = \frac{\lambda_{\mathrm{rep}}}{\lambda_{\mathrm{rep}} - \lambda_{\mathrm{att}}}\left(\frac{\lambda_{\mathrm{rep}}}{\lambda_{\mathrm{att}}}\right)^{\frac{\lambda_{\mathrm{att}}}{\lambda_{\mathrm{rep}} - \lambda_{\mathrm{att}}}}, \tag{2.99}$$

where U^{Mie} is the Mie potential energy, r is the distance between the centers of two segments, ε is the depth of the potential energy well, σ is the distance at which the potential is zero, λ_{att} is the attractive range potential (usually the value is set to 6), and λ_{rep} is the repulsive range potential (therefore, an adjustable parameter).

In SAFT-VR Mie, each molecule is formed by homonuclear segments (*i.e.*, each chemically distinct group is represented by segments of the same type). In another existing approach in the literature proposed by Papaioannou *et al.* [88], each molecule is formed by heteronuclear segments (*i.e.*, segments of different types represent each chemically distinct group). The variant in which SAFT-VR Mie is extended to a heteronuclear formulation is called SAFT-γ Mie [88]. As SAFT is a molecular-based, equation of state, the parameters fitted within SAFT-γ Mie can be used as a coarse-grained force field for molecular simulations. In the literature [88], the coarse-grained force fields obtained by SAFT-γ Mie are referred to as SAFT-γ Mie force fields.

2.1.8.2 SAFT-VR Mie equation of state

The SAFT-VR Mie equation of state is one of the last versions of SAFT theory. The SAFT-VR Mie equation of state was originally proposed by Lafitte *et al.* [85,86]. The first works

using the SAFT-VR Mie were published in 2006 [85] and 2007 [86], with an improved version of this model presented in 2013 [87]. In this Section, only the mathematical expressions of the main contributions in the SAFT-VR Mie equation of state are presented, with the description of the improved model available in Appendix B (in a simplified version) or in the original publication [87] (in the complete version). The mathematical expressions presented in this section are valid for mixtures.

In the SAFT-VR Mie equation of state, the interaction of the fluids is modeled by the Mie potential, which can be described for a mixture of monomers (segments) of species i and j according to Eqs. (2.100) and (2.101):

$$U_{ij}^{\text{Mie}}(r) = C_{ij}\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{\lambda_{\text{rep},ij}} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{\lambda_{\text{att},ij}}\right], \tag{2.100}$$

$$C_{ij} = \frac{\lambda_{\text{rep},ij}}{\lambda_{\text{rep},ij} - \lambda_{\text{att},ij}}\left(\frac{\lambda_{\text{rep},ij}}{\lambda_{\text{att},ij}}\right)^{\frac{\lambda_{\text{att},ij}}{\lambda_{\text{rep},ij} - \lambda_{\text{att},ij}}}, \tag{2.101}$$

where U^{Mie} is the Mie potential energy relative to the monomers i and j, r_{ij} is the relative distance between the centers of monomers i and j, σ is the distance between monomers i and j in which the Mie potential is zero, ε is the depth of the potential energy well relative to the monomers i and j, λ^{rep} is the repulsive range potential relative to the monomers i and j (it is an adjustable parameter), and λ^{att} is the attractive range potential relative to the monomers i and j (usually set equal to 6, since both the Boltzmann-averaged, classical dipole-dipole interactions and the leading-order, quantum-mechanical dispersion interactions are proportional to $-1/r_{ij}^6$).

For mixture calculations, pure component parameters should be used to compute mixture-specific parameters. In SAFT-VR Mie, these mixture-specific parameters for a mixture of monomers of species i and j are obtained by combining rules given by the set of Eqs. (2.102)-(2.104):

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2}, \tag{2.102}$$

$$\varepsilon_{ij} = \frac{\sqrt{\sigma_{ii}^3 \sigma_{jj}^3}}{\sigma_{ij}^3}\sqrt{\varepsilon_{ii}\varepsilon_{jj}}, \tag{2.103}$$

$$\lambda_{k,ij} - 3 = \sqrt{\left(\lambda_{k,ii} - 3\right)\left(\lambda_{k,jj} - 3\right)} \quad \text{for } k = \text{att, rep.} \tag{2.104}$$

In the SAFT-VR Mie equation of state, the total dimensionless Helmholtz free energy of a fluid of associating chain molecules formed from Mie segments can be described in terms of ideal gas, monomer, segment, chain, and associative contributions by Eq. (2.105):

$$a^{\text{total}} = a^{\text{ideal}} + a^{\text{mono}} + a^{\text{chain}} + a^{\text{assoc}}, \tag{2.105}$$

where a is the dimensionless Helmholtz free energy[20], and superscripts total, ideal, mono, chain, and assoc are related to the total, ideal gas, monomer, segment, chain, and associating contributions, respectively.

The ideal gas, monomer, segment, chain, and associating contributions to the Helmholtz free energy can be expressed in a mixture of "i" species by Eqs. (2.106), (2.107), (2.108), and (2.109), respectively:

$$a^{\text{ideal}} = \left(\sum_{i=1}^{n} x_i \ln \rho_i \Lambda_i^3 \right) - 1, \tag{2.106}$$

$$a^{\text{mono}} = \left(\sum_{i=1}^{n} x_i m_i \right) a^{\text{M}}, \tag{2.107}$$

$$a^{\text{chain}} = -\sum_{i=1}^{n} x_i \left(m_i - 1\right) \ln g_{ii}^{\text{Mie}} \left(\sigma_{ii}\right), \tag{2.108}$$

$$a^{\text{assoc}} = \sum_{i=1}^{n} x_i \sum_{a=1}^{s_i} n_{ai} \left[\ln X_{ai} - \frac{X_{ai}}{2} + \frac{1}{2} \right], \tag{2.109}$$

where a is the dimensionless Helmholtz free energy, n is the number of species in the mixture, x_i is the mole fraction of the specie i in the mixture[21], ρ_i is the density of specie i in the mixture[22], Λ_i is the de Broglie thermal wavelength of species i in the mixture, m_i is the number of monomers in the chain i, a^{M} is the monomer Helmholtz free energy[23], g_{ii}^{Mie} is an analytical expression for radial distribution function, s_i is number of association sites of species i in the mixture, n_{ai} is the number of sites of type a in species i in the mixture, and X_{ai} is the fraction of species i in the mixture that is not bonded to the site of type a.

2.2 Computational details

2.2.1 General scheme

Fig. 2 presents a general representative scheme that comprises the main steps relating to calculations performed in this study. As shown in Fig. 2, this study comprised two main steps:

[20] The dimensionless Helmholtz free energy is given by $a = A/Nk_BT$, where a is the dimensionless Helmholtz free energy, A is the Helmholtz free energy, N is the total number of segment units, k_{B} is the Boltzmann constant, T is the absolute temperature.

[21] The mole fraction is given by $x_i = N_i/N$, where N_i is the number of molecules of specie i in the mixture and N is the total number of molecules.

[22] The density is given by $\rho_i = N_i/V$, where N_i is the number of molecules of specie i in the mixture, and V is the total volume of the mixture.

[23] The monomer Helmholtz free energy is expressed as a third-order expansion given by $a^{\text{M}} = a^{\text{HS}} + \beta a_1 + \beta^2 a_2 + \beta^3 a_3$, where a^{HS} is an expression for a multicomponent mixture of hard spheres, $\beta = 1/(k_{\text{B}}T)$ (with k_{B} the Boltzmann constant and T the absolute temperature, and a_1, a_2, and a_3 are perturbation terms of first, second, and third-order, respectively

(i) obtaining the Mie potential parameters through *ab initio* calculations and (ii) evaluating the Mie potential parameters obtained by *ab initio* calculations in the prediction of experimental data through the SAFT-VR Mie equation of state and molecular dynamics simulations.

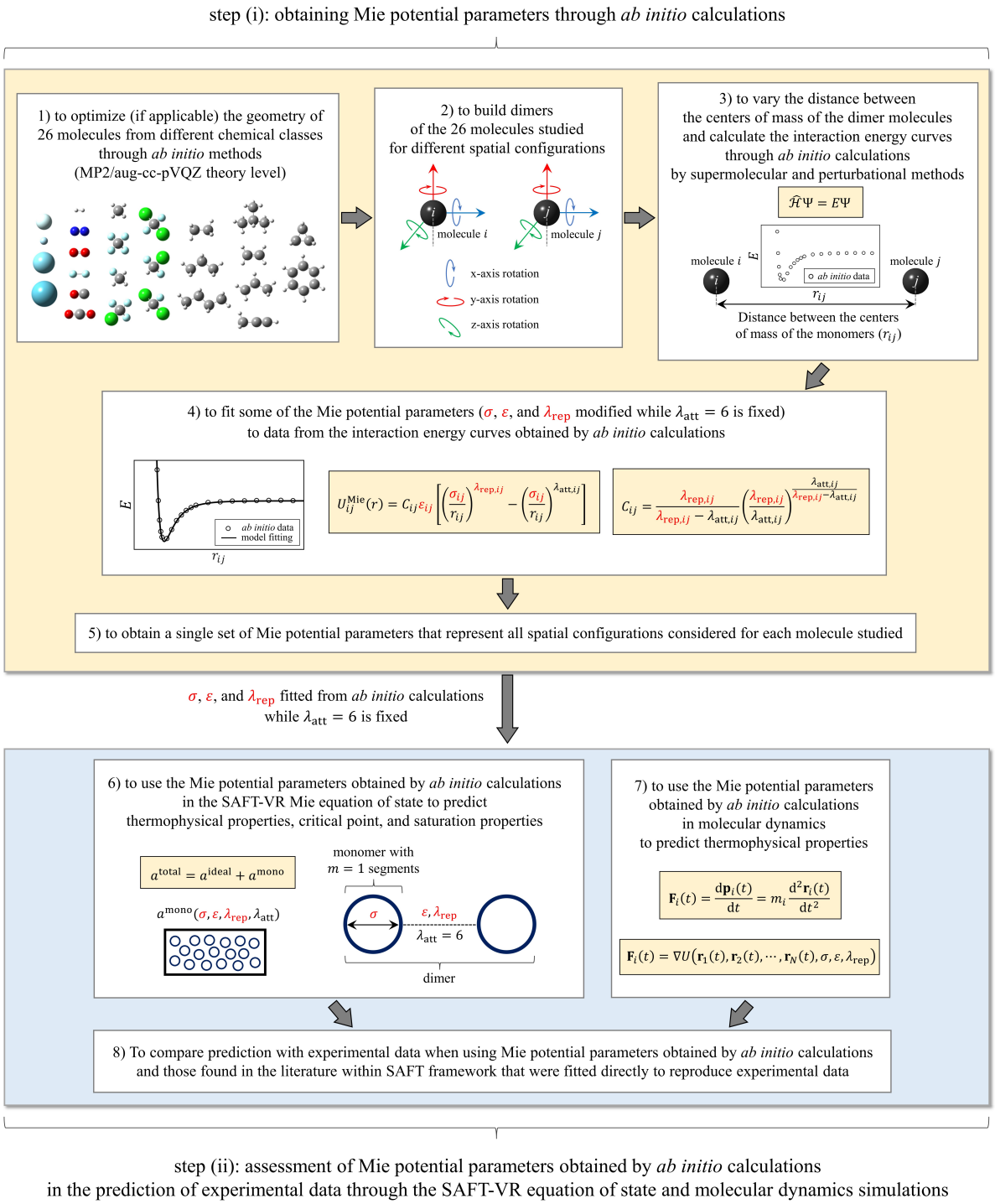

Figure 2 – General representative scheme of the main calculation steps performed in this study.

In step (i), a group of 26 molecules from different chemical classes was first selected. The geometry of the 26 molecules was optimized using *ab initio* methods. Molecules with optimized geometries were used to build dimers of different spatial configurations, rotating each

molecule of the dimer in the three possible Cartesian axes. The distances between the centers of mass of the dimer molecules of different spatial configurations were varied. Interaction energy curves were constructed through *ab initio* calculations using supermolecular and perturbational methods. The Mie potential parameters σ, ε, and λ_{rep} were fitted to the *ab initio* data while λ_{att} was fixed equal to 6. A single set of Mie potential parameters σ, ε, λ_{rep}, and λ_{att} was obtained for each of the 26 molecules studied, considering these as a spherical 1-site coarse-grained model.

In step (ii), the Mie potential parameters obtained by *ab initio* calculations were used in the SAFT-VR Mie equation of state to calculate thermophysical properties, critical point and saturation properties. Only the ideal gas and monomer contributions were considered in the version of the SAFT-VR Mie equation of state used in this study (see Eq. (2.105) for details). The dependence of monomer contribution in relation to the Mie potential parameters can be verified in the development of the equations presented in Appendix B. The thermophysical properties were also evaluated for some specific molecules by molecular dynamics simulations. Finally, the prediction capabilities of experimental data were compared when using Mie potential parameters obtained by *ab initio* calculations and using those found in the literature that were directly fitted to reproduce experimental data.

More details about the steps described in this Section are presented later in this dissertation. In section 2.2.2, details of step (i) are presented, while in Sections 2.2.3 and 2.2.4, more details regarding step (ii) are presented.

2.2.2 Obtaining Mie potential parameters through *ab initio* calculations

In this study, Gaussian 09 [90] and psi4 [91] software were used to perform all *ab initio* calculations. The bottom-up methodology was considered to obtain Mie potential parameters for 26 molecules from various chemical classes. This list includes light gases (helium, neon, argon, and krypton, hydrogen, nitrogen, oxygen, fluorine, carbon monoxide, and carbon dioxide), linear hydrocarbons (methane, ethane, propane, and *n*-butane), branched hydrocarbons (isobutane), unsaturated hydrocarbons (ethene, propene, and propyne), cyclic hydrocarbons (cyclopropane), aromatic hydrocarbons (benzene), and refrigerants (tetrafluoromethane, fluoromethane, chlorotrifluoromethane, dichlorodifluoromethane, dichlorofluoromethane, and chlorodifluoromethane).

First, when applicable, the equilibrium geometry of the studied molecule was obtained at a high level of theory (in this study, it was MP2/aug-cc-pVQZ theory level). An example script for this task is the code 1 available in the Appendix A. To obtain interaction energy curves through *ab initio* calculations, dimers of different conformations were built using molecules of the same type with equilibrium geometry. The studied molecules were divided into three main groups to obtain the dimers, according to Fig. 3. The first group included some noble gases (*i.e.*, helium, neon, argon, and krypton). The second group consisted of methane and

some substituted-methane compounds (*i.e.*, methane and refrigerants). Finally, the third group contained the remaining molecules from the set of 26.

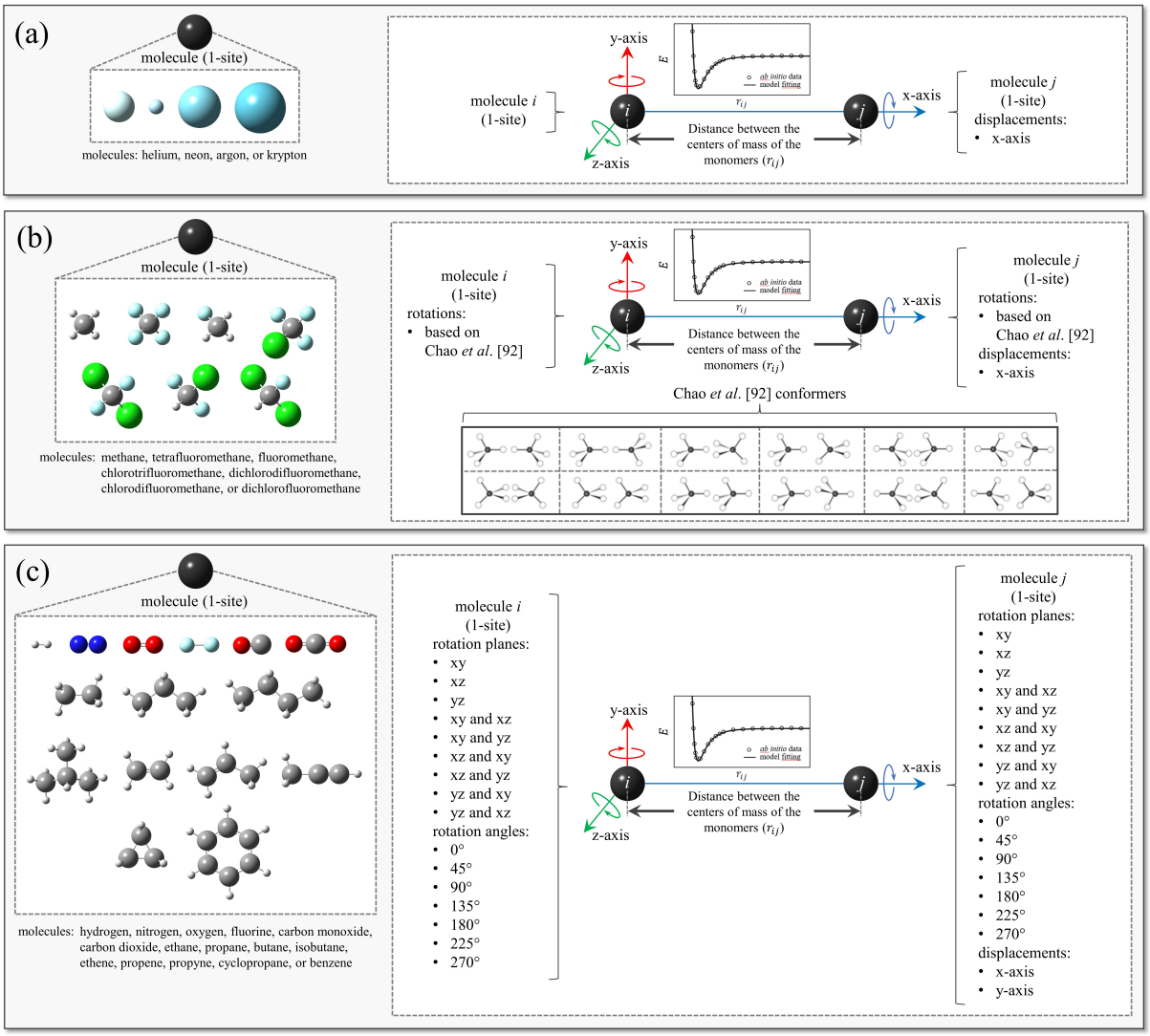

Figure 3 – Representative scheme for constructing dimers to obtain interaction energy curves through *ab initio* calculations. Illustration of the procedure performed for (a) noble gases, (b) methane and substituted-methane compounds, and (c) other molecules considered in this work (*i.e.*, molecules of different chemical classes).

In the first group of molecules, interaction energy curves were obtained by varying the distance of the dimer molecules of some noble gases in one of the Cartesian reference axes, as illustrated in Fig. 3(a). In the second group of molecules, interaction energy curves were obtained by varying the distance from the centers of mass of the dimer molecules in one of the Cartesian reference axes. Furthermore, in the second group of molecules, rotations were applied to each molecule of the dimer to produce 12 fundamental symmetric conformations according to Chao *et al.* [92], as depicted in Fig. 3(b). Additionally, when convenient, the positions of different atoms bonded to the carbon atom were also permuted in the configurations proposed by Chao *et al.* [92].

In the third group of molecules, interaction energy curves were obtained by varying the distance from the centers of mass of the dimer molecules in two of the Cartesian reference axes. Furthermore, each molecule of the dimer was rotated at different angles in one or two planes. The Cartesian reference axes in which the distance from the centers of mass of the dimer molecules varied and the angles and planes of rotation applied to each molecule are illustrated in Fig. 3(c). Repeated configurations obtained in this process were discarded.

Interaction energy curves as a function of the distance from the centers of mass of the dimer molecules were obtained by supermolecular and perturbational methods.[24]. Examples of scripts for this task are the codes 2 and 3 available in the Appendix A. By supermolecular methods, energy calculations were performed at counterpoise-corrected second and fourth-order Møller-Plesset perturbation theory (*i.e.,* MP2 and MP4 energies, respectively) as well as at coupled-cluster with singles, doubles, and perturbative triples excitations (*i.e.,* CCSD(T) energies). The energies of the supermolecular methods were obtained using Dunning basis sets. By perturbational methods, energy calculations were performed through symmetry-adapted perturbation theory (SAPT) considering bronze, silver, and gold standards (*i.e.*, sSAPT0/jun-cc-pVDZ, SAPT2+/aug-cc-pVDZ, and SAPT2+(3)δMP2/aug-cc-pVTZ energies, respectively).

The studied molecules were represented using a spherical 1-site coarse-grained model with interaction based on the Mie potential. The Mie potential parameters σ, ε, and λ^{rep} were fitted to represent interaction energy data from *ab initio* calculations. The parameter λ^{att} of the Mie potential was set to 6, similar to most related SAFT models. A single set of Mie potential parameters was fitted for each studied molecule to represent the interaction energy data for various dimer configurations. In this task, two parameter fitting approaches were investigated.

In the first approach (called approach A), a set of Mie potential parameters was fitted to represent the interaction energies for each dimer configuration of the studied molecule. A unique set of Mie potential parameters was obtained as an average of the set of parameters for each dimer configuration of the molecule. This approach was not considered in the sole reference on this topic found in the literature [10]. In approach A, the objective function that must be minimized to obtain a unique set of Mie potential parameters is given by Eq. (2.110):

$$\min_{\Omega_i} F_{\mathrm{obj}}^{\mathrm{A}} = \sum_{j=1}^{N^{\mathrm{dist}}} \left[\frac{E^{\mathrm{QM}}\left(r_j, w_i\right) - E^{\mathrm{Mie}}\left(r_j, \Omega_i\right)}{E^{\mathrm{QM}}\left(r_j, w_i\right)} \right]^2, \tag{2.110}$$

$$\left\langle \Omega_k \right\rangle = \frac{1}{N^{\mathrm{conf}}} \sum_{i=1}^{N^{\mathrm{conf}}} \Omega_{ik}, \tag{2.111}$$

where $F_{\mathrm{obj}}^{\mathrm{A}}$ is the objective function to be minimized for approach A, Ω_i represents the Mie potential parameters estimated for the i conformer ($\Omega_i = [\sigma_i, \varepsilon_i, \lambda_i^{\mathrm{rep}}]$), N^{dist} is the number of

[24] The geometry of the molecules was optimized at a single level of theory (*i.e.*, in this work at MP2/aug-cc-pVQZ theory level) to make a coherent comparison of the calculation of the interaction energy of dimers using different methods (*i.e.*, supermolecular and perturbational)

distances evaluated between the dimers of the i conformer, E^{QM} is the *ab initio* interaction energy, E^{Mie} is the Mie potential energy, r_j is the j^{th} point considered for distance between the dimers of the i configuration, w_i represents the spatial orientation of the i conformer, $\Omega = [\langle\sigma_i\rangle, \langle\varepsilon_i\rangle, \langle\lambda_i^{\mathrm{rep}}\rangle]$ represents the final set of the Mie potential parameters, and N^{conf} is the number of evaluated conformers.

In the second approach (called approach B), similar to the sole reference on this topic found in the literature [10], a unique set of Mie potential parameters was obtained considering in the parameter-fitting procedure all interaction energies of all configurations at once for each studied molecule. For this approach, the deviations between the interaction energies obtained from *ab initio* calculations and those predicted with the Mie potential were weighted according to the Boltzmann energy distribution function. In approach B, the objective function that must be minimized to obtain a unique set of Mie potential parameters is given by Eq. (2.112):

$$\min_{\Omega} F_{\mathrm{obj}}^{\mathrm{B}} = \sum_{i=1}^{N^{\mathrm{conf}}} \sum_{j=1}^{N^{\mathrm{dist}}} \left[\frac{E^{\mathrm{QM}}\left(r_j, w_i\right) - E^{\mathrm{Mie}}\left(r_j, \Omega\right)}{E^{\mathrm{QM}}\left(r_j, w_i\right)} \right]^2 \phi\left(r_j, w_i\right), \tag{2.112}$$

$$\phi\left(r_j, w_i\right) = \frac{e^{\left[-E^{\mathrm{QM}}\left(r_j, w_i\right)/(k_{\mathrm{B}}T)\right]}}{\sum_{i=1}^{N^{\mathrm{conf}}} \sum_{j=1}^{N^{\mathrm{dist}}} e^{\left[-E^{\mathrm{QM}}\left(r_j, w_i\right)/(k_{\mathrm{B}}T)\right]}}, \tag{2.113}$$

where $F_{\mathrm{obj}}^{\mathrm{B}}$ is the objective function to be minimized for approach B, $\Omega = [\sigma, \varepsilon, \lambda^{\mathrm{rep}}]$ represents the set of the estimated Mie potential parameters, N^{conf} is the number of conformers evaluated, N^{dist} is the number of distances evaluated between the dimers in the conformers, r_j is the j^{th} point considered for distance considered between the dimers in the interaction energy curve of the conformers, w_i represents the spatial orientation of the i conformer, E^{QM} is the *ab initio* energy, E^{Mie} is the Mie potential energy, ϕ is a weighting factor[25], k_{B} is the Boltzmann constant, and T is the absolute temperature (here adopted as 273.15 K, which is an arbitrary value)[26].

To estimate the Mie potential parameters from interaction energy data obtained through *ab initio* calculations, the Levenberg-Marquardt algorithm [93, 94] was used in the procedure. To avoid the local minimum of the objective functions described by Eqs. (2.110) and (2.112), a multi-start parameter fitting strategy was considered. The set of Mie potential parameters obtained for each molecule studied through *ab initio* calculations (*i.e.*, bottom-up methodology) was compared with those from literature within the SAFT framework context (*i.e.*, top-down methodology).

Thermophysical properties (*e.g.*, density, isochoric heat capacity, isobaric heat capacity, and speed of sound) were computed using top-down and bottom-up-based Mie potential

[25] $\sum_{i=1}^{N^{\mathrm{conf}}} \sum_{j=1}^{N^{\mathrm{dist}}} \phi\left(r_j, w_i\right) = 1$

[26] The impact of small changes in temperature causes negligible variations in the Boltzmann weight: going from 273.15 K to 298.15 K would imply a difference in the ten thousandth place in the weighting factor.

parameters in the SAFT-VR Mie equation of state and through molecular dynamics simulations. Critical point (*i.e.*, critical temperature, pressure, and density) and saturation properties (*e.g.*, saturation pressure and enthalpy of vaporization) were computed using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state. The thermophysical properties and critical point obtained using top-down and bottom-up-based Mie potential parameters were compared with the National Institute of Standards and Technology (NIST) reference data [95]. The comparison made for the critical point was through the percentage deviation, and for the thermophysical and saturation properties was through the average absolute relative deviation (AARD%), given by Eq.(2.114):

$$\text{AARD\%} = \frac{1}{N_{\text{data}}} \sum_{i=1}^{N_{\text{data}}} \left| \frac{y_i^{\text{ref}} - y_i^{\text{sim}}}{y_i^{\text{ref}}} \right| \times 100, \tag{2.114}$$

where AARD% is the average absolute relative deviation, N_{data} is the number of reference data points, y_i^{ref} is the thermodynamic property at the conditions of the i^{th} reference point, y_i^{sim} is the thermodynamic property obtained via molecular dynamics simulations and SAFT-VR Mie equation of state at the conditions of the i^{th} reference point.

2.2.3 Calculation of thermophysical properties by molecular dynamics simulations

In this study, LAMMPS [96] software was used to perform molecular dynamics simulations for some of the systems studied. The simulated systems consisted of 1000 molecules arranged in a cubic simulation box, with initial configurations built using Moltemplate [97] and Packmol [98] software. Periodic boundary conditions were applied to all directions of the simulation box. The molecular dynamics simulations had three steps: energy minimization, equilibration, and production. The energy minimization step used the Polak-Ribiere version of the conjugate gradient algorithm. During the equilibration step, the studied systems were simulated for 2 ns in the canonical ensemble (NVT), followed by another 3 ns in the isothermal-isobaric ensemble (NPT). For the production step, the studied systems were simulated for 5 ns in the isothermal-isobaric (NPT) ensemble.

In the molecular dynamics simulations, the equations of motion were integrated using the velocity-Verlet algorithm with a 1 fs integration step. The Mie potential had a cutoff of 14 Å, and tail corrections were added to both energy and pressure. To control temperature and pressure, the Nosé-Hoover thermostat and barostat were used with damping parameters of 100 fs and 1000 fs, respectively.

The top-down and bottom-up-based Mie potential parameters were used in molecular dynamics simulations to compute thermophysical properties such as density, isochoric and isobaric heat capacities, and the speed of sound. Derivative properties were computed during the

production step in the isothermal-isobaric (NPT) ensemble following relations from the set of Eqs. (2.115)-(2.119) derived from fluctuation theory [99]:

$$\alpha = N\left[\frac{\left\langle vh^{\text{conf}}\right\rangle - \langle v\rangle\left\langle h^{\text{conf}}\right\rangle}{\langle v\rangle RT^2}\right], \tag{2.115}$$

$$k_{\text{T}} = N\left[\frac{\left\langle v^2\right\rangle - \langle v\rangle^2}{\langle v\rangle RT^2}\right], \tag{2.116}$$

$$c_{\text{p}} = \sum_{i=0}^{4} a_i RT^i + N\frac{\left\langle u^{\text{conf}}h^{\text{conf}}\right\rangle - \left\langle u^{\text{conf}}\right\rangle\left\langle h^{\text{conf}}\right\rangle}{RT^2} + p\langle v\rangle\alpha - R, \tag{2.117}$$

$$c_{\text{v}} = c_{\text{p}} - \frac{T\langle v\rangle\alpha^2}{k_T}, \tag{2.118}$$

$$c_{\text{sound}} = \sqrt{\frac{c_{\text{p}}}{c_{\text{v}}k_{\text{T}}\rho}}, \tag{2.119}$$

where α is the thermal expansion coefficient, N is the number of molecules, v is the simulation box volume, h^{conf} is the configurational enthalpy (which takes into account only the intermolecular contributions), R is the universal gas constant, T is the absolute temperature, k_{T} is the isothermal compressibility, c_{p} is the isobaric heat capacity, a_i is the i^{th} term for the equation of ideal gas isobaric heat capacity polynomial function, whose values were obtained in the book of Poling *et al.* [100], u^{conf} is the configurational internal energy (which takes into account only intermolecular contributions), p is the pressure, c_{v} is the isochoric heat capacity, c_{sound} is the speed of sound, ρ is the density obtained directly from the simulation, and $\langle\rangle$ represent ensemble averages.

The block average approach was used to compute the uncertainties, estimating the statistical inefficiency from linear regression between 2 and 100 blocks [101]. The covariance matrix was used to propagate errors for derivative properties [101].

2.2.4 Calculation of thermophysical properties, critical point, and saturation properties with SAFT-VR Mie equation of state

The SAFT-VR Mie equation of state was implemented in an in-house code of the Laboratory of Complex Systems Engineering at the University of Campinas (LESC/Unicamp). The SAFT-VR Mie code was developed in the Fortran 95 programming language. The SAFT-VR Mie code consists of the mathematical extension of the ideal gas, monomer, segmentation, chain, and associating contributions, where additional equations can be found in the original publication by Lafitte *et al.* [87].

The top-down and bottom-up-based Mie potential parameters were used in the SAFT-VR Mie equation of state to compute thermophysical properties such as density, isochoric and isobaric heat capacities, the speed of sound, and the Joule-Thomson coefficient. Similar to

molecular dynamics simulations, derivative properties were also computed with the SAFT-VR Mie equation of state. Derivative properties from SAFT-VR Mie equation of state can be obtained as a function of free Helmholtz energy and pressure according to the set of Eqs. (2.120)-(2.124) [102]:

$$c_{\mathrm{v}} = \sum_{i=0}^{4} a_i R T^i - R - T\left(\frac{\partial^2 A^{\mathrm{res}}}{\partial T^2}\right)_v, \tag{2.120}$$

$$k_{\mathrm{T}}^{-1} = \rho\left(\frac{\partial p}{\partial \rho}\right)_T, \tag{2.121}$$

$$\alpha = k_{\mathrm{T}}\left(\frac{\partial p}{\partial T}\right)_v, \tag{2.122}$$

$$c_{\mathrm{p}} = c_{\mathrm{v}} + \frac{T\alpha^2}{k_T}, \tag{2.123}$$

$$c_{\mathrm{sound}} = \sqrt{\frac{c_{\mathrm{p}}}{c_{\mathrm{v}}}\left(\frac{\partial p}{\partial \rho}\right)_T}, \tag{2.124}$$

$$\mu_{\mathrm{JT}} = T\left(\frac{\partial p}{\partial T}\right)_v - \rho\left(\frac{\partial p}{\partial \rho}\right)_T, \tag{2.125}$$

where c_{v} is the isochoric heat capacity, a_i is the i^{th} term for the equation of ideal gas isobaric heat capacity polynomial function, whose values were obtained in the book of Poling *et al.* [100], R is the universal gas constant, T is the the absolute temperature, A^{res} is the residual Helmholtz free energy, k_{T} is the isothermal compressibility, ρ is the density, p is the pressure, α is the thermal expansion coefficient, v is the volume, c_{p} is the isobaric heat capacity, c_{sound} is the speed of sound, and μ_{JT} is the Joule-Thomson coefficient.

The critical point can be obtained by the SAFT-VR Mie equation of state by simultaneously solving the system of equations for pressure given by Eqs. (2.126) and (2.127)

$$\left(\frac{\partial p}{\partial \rho}\right)_T = 0, \tag{2.126}$$

$$\left(\frac{\partial^2 p}{\partial \rho^2}\right)_T = 0. \tag{2.127}$$

The saturation properties can also be calculated by the SAFT-VR Mie equation of state by determining the vapor and liquid densities that satisfy, for a given temperature, the equalities for pressure given by Eq. (2.128) and for the chemical potential given by Eq. (2.129):

$$p = -\left(\frac{\partial A^{\mathrm{res}}}{\partial v}\right)_{N,T}, \tag{2.128}$$

$$\mu = \left(\frac{\partial A}{\partial N}\right)_{v,T}, \tag{2.129}$$

where N is the number of molecules, and μ is the chemical potential,

3 Results and discussion

"The laws of thermodynamics, as empirically determined, express the approximate and probable behavior of systems of a great number of particles, or, more precisely, they express the laws of mechanics for such systems as they appear to beings who have no the fineness of perception to enable them to appreciate quantities of the order of magnitude of those which relate to single particles, and who cannot repeat their experiments often enough to obtain any but the most probable results."
Josiah Willard Gibbs

3.1 First group: noble gases

Interaction energy curves were obtained by supermolecular and perturbational methods varying the distance of the dimer molecules of some noble gases, according to the procedure previously described in Fig. 3(a) in Section 2.2.2 of this study. The supermolecular calculations involved obtaining the second and fourth-order Møller-Plesset energies (*i.e.*, MP2 and MP4 energies, respectively) and coupled-cluster with singles, doubles, and perturbative triples excitations energies (*i.e.*, CCSD(T) energy). In supermolecular calculations, double, triple, and quadruple-zeta Dunning basis sets were used (*i.e.*, aug-cc-pV*n*Z, with n = D, T, and Q, respectively). The perturbational calculations involved obtaining symmetry-adapted perturbation theory (SAPT) energies at bronze, silver, and gold standard levels (*i.e.*, sSAPT0/jun-cc-pVDZ, SAPT2+/aug-cc-pVDZ, and SAPT2+(3)δMP2 /aug-cc-pVTZ energies, respectively).

The Mie potential parameters σ, ε, and λ^{rep} were fitted to interaction energy curves obtained through *ab initio* calculations by a nonlinear regression procedure. Since noble gases are monatomic, the objective function given by Eq. (2.110) presented in Section 2.2.2 of this study for approach A considering a single configuration was minimized in the parameter fitting procedure.

Table 1 presents the Mie potential parameters obtained through the top-down and bottom-up based methodologies considered in this study. Top-down-based Mie potential parameters were obtained in the literature [103, 104] within the SAFT framework (*i.e.* SAFT-VR Mie equation of state). Bottom-up-based Mie potential parameters denote the parameters fitted to represent interaction energy curves obtained through *ab initio* calculations. The only available reference for bottom-up-based Mie potential parameters is from Walker *et al.* [10]. The Mie potential energy curves obtained in this study using parameters from the bottom-up methodology and the interaction energy curves obtained through *ab initio* calculations by supermolecular and perturbational methods are presented in Fig. 4. An analysis of Table 1 and Fig. 4 reveals some patterns of results.

Table 1 – Top-down and bottom-up-based Mie potential parameters for some noble gases obtained by different methodologies and contexts.

			Mie potential parameters[a,b]					
Molecule	Methodology[c]	Context[d]	m	σ/Å	λ_{rep}	λ_{att}	(ε/k_B)/K	Reference
helium	TD	SAFT framework	1.0000	3.3530	14.8400	6.0000	4.4400	[104]
	BU (S)	DLPNO-CCSD(T)/5Z	1.0000	2.6876	10.6560	6.0000	9.5468	[10]
		MP2/aug-cc-pVDZ (A)	1.0000	2.8859	8.9532	6.0000	6.6381	this work
		MP2/aug-cc-pVTZ (A)	1.0000	2.8475	9.2472	6.0000	6.7890	this work
		MP2/aug-cc-pVQZ (A)	1.0000	2.8293	9.7334	6.0000	6.9399	this work
		MP4/aug-cc-pVDZ (A)	1.0000	2.8431	12.3129	6.0000	5.5318	this work
		MP4/aug-cc-pVTZ (A)	1.0000	2.7554	13.3430	6.0000	7.6942	this work
		MP4/aug-cc-pVQZ (A)	1.0000	2.7378	12.9627	6.0000	8.2474	this work
		CCSD(T)/aug-cc-pVDZ (A)	1.0000	2.8379	12.3241	6.0000	5.6826	this work
		CCSD(T)/aug-cc-pVTZ (A)	1.0000	2.7226	14.8101	6.0000	8.8508	this work
		CCSD(T)/aug-cc-pVQZ (A)	1.0000	2.7324	12.9737	6.0000	8.5491	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	1.0000	3.3177	11.9702	6.0000	0.8549	this work
		SAPT2+/aug-cc-pVDZ (A)	1.0000	2.8771	11.0784	6.0000	7.8451	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	1.0000	2.7788	10.3647	6.0000	9.1526	this work
neon	TD	SAFT framework	1.0000	2.8019	9.6977	6.0000	29.8750	[103]
	BU (S)	DLPNO-CCSD(T)/5Z	1.0000	2.8249	10.648	6.0000	35.7160	[10]
		MP2/aug-cc-pVDZ (A)	1.0000	3.0825	13.1555	6.0000	10.8624	this work
		MP2/aug-cc-pVTZ (A)	1.0000	2.9875	11.4790	6.0000	17.1988	this work
		MP2/aug-cc-pVQZ (A)	1.0000	2.9148	12.2795	6.0000	22.0265	this work
		MP4/aug-cc-pVDZ (A)	1.0000	3.0563	12.6671	6.0000	13.1254	this work
		MP4/aug-cc-pVTZ (A)	1.0000	2.8835	11.7266	6.0000	24.4906	this work
		MP4/aug-cc-pVQZ (A)	1.0000	2.8315	13.0187	6.0000	31.5311	this work
		CCSD(T)/aug-cc-pVDZ (A)	1.0000	3.1182	12.7248	6.0000	11.7173	this work
		CCSD(T)/aug-cc-pVTZ (A)	1.0000	2.8619	12.0503	6.0000	25.2953	this work
		CCSD(T)/aug-cc-pVQZ (A)	1.0000	2.8273	13.4409	6.0000	32.2351	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	1.0000	3.2139	13.1770	6.0000	7.5936	this work
		SAPT2+/aug-cc-pVDZ (A)	1.0000	3.0432	12.7364	6.0000	14.0306	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	1.0000	2.8882	11.7386	6.0000	24.4404	this work
argon	TD	SAFT framework	1.0000	3.4038	12.0850	6.0000	117.8400	[103]
		SAFT framework	1.0000	3.4140	14.8500	6.0000	132.0400	[104]
	BU (S)	DLPNO-CCSD(T)/5Z	1.0000	3.4298	11.8210	6.0000	120.5300	[10]
		MP2/aug-cc-pVDZ (A)	1.0000	3.5705	13.1696	6.0000	78.4505	this work
		MP2/aug-cc-pVTZ (A)	1.0000	3.4295	13.5842	6.0000	121.8498	this work
		MP2/aug-cc-pVQZ (A)	1.0000	3.3636	13.1097	6.0000	141.5630	this work
		MP4/aug-cc-pVDZ (A)	1.0000	3.6316	13.4156	6.0000	62.6599	this work
		MP4/aug-cc-pVTZ (A)	1.0000	3.4710	13.7601	6.0000	104.9527	this work
		MP4/aug-cc-pVQZ (A)	1.0000	3.4004	13.2812	6.0000	123.8110	this work
		CCSD(T)/aug-cc-pVDZ (A)	1.0000	3.6350	13.4160	6.0000	61.9055	this work
		CCSD(T)/aug-cc-pVTZ (A)	1.0000	3.4790	13.7676	6.0000	101.8348	this work
		CCSD(T)/aug-cc-pVQZ (A)	1.0000	3.4116	13.2805	6.0000	118.7319	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	1.0000	4.1706	16.1386	6.0000	13.1254	this work
		SAPT2+/aug-cc-pVDZ (A)	1.0000	3.6067	13.3678	6.0000	65.7275	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	1.0000	3.4757	13.7510	6.0000	103.7961	this work
krypton	TD	SAFT framework	1.0000	3.6359	11.9850	6.0000	163.2700	[103]
	BU (S)	DLPNO-CCSD(T)/5Z	1.0000	3.6572	12.5720	6.0000	168.0600	[10]
		MP2/aug-cc-pVDZ (A)	1.0000	3.8386	13.1055	6.0000	101.7845	this work
		MP2/aug-cc-pVTZ (A)	1.0000	3.6657	13.1544	6.0000	175.0051	this work
		MP2/aug-cc-pVQZ (A)	1.0000	3.5804	13.1776	6.0000	215.7390	this work
		MP4/aug-cc-pVDZ (A)	1.0000	3.9266	13.5295	6.0000	77.4951	this work
		MP4/aug-cc-pVTZ (A)	1.0000	3.7308	13.8585	6.0000	146.2399	this work
		MP4/aug-cc-pVQZ (A)	1.0000	3.6327	13.7912	6.0000	184.3085	this work
		CCSD(T)/aug-cc-pVDZ (A)	1.0000	3.9334	13.5051	6.0000	75.5338	this work
		CCSD(T)/aug-cc-pVTZ (A)	1.0000	3.7481	13.8716	6.0000	137.4393	this work
		CCSD(T)/aug-cc-pVQZ (A)	1.0000	3.6558	13.8100	6.0000	169.3224	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	1.0000	4.2524	12.9636	6.0000	25.6976	this work
		SAPT2+/aug-cc-pVDZ (A)	1.0000	3.9053	13.4105	6.0000	79.6072	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	1.0000	3.7343	13.8864	6.0000	144.7815	this work

[a]As helium, neon, argon, and krypton are monatomic, these molecules are modeled as a spherical 1-site (in the SAFT-VR Mie equation of state is equivalent to assigning the number of segments m equal to 1). [b]k_B is the Boltzmann constant. [c]TD and BU refer to the top-down and bottom-up methodologies, respectively, while (S) and (P) refer to the supermolecular and perturbational methods, respectively. [d](A) indicates that the objective function of approach A was minimized in the parameter fitting procedure.

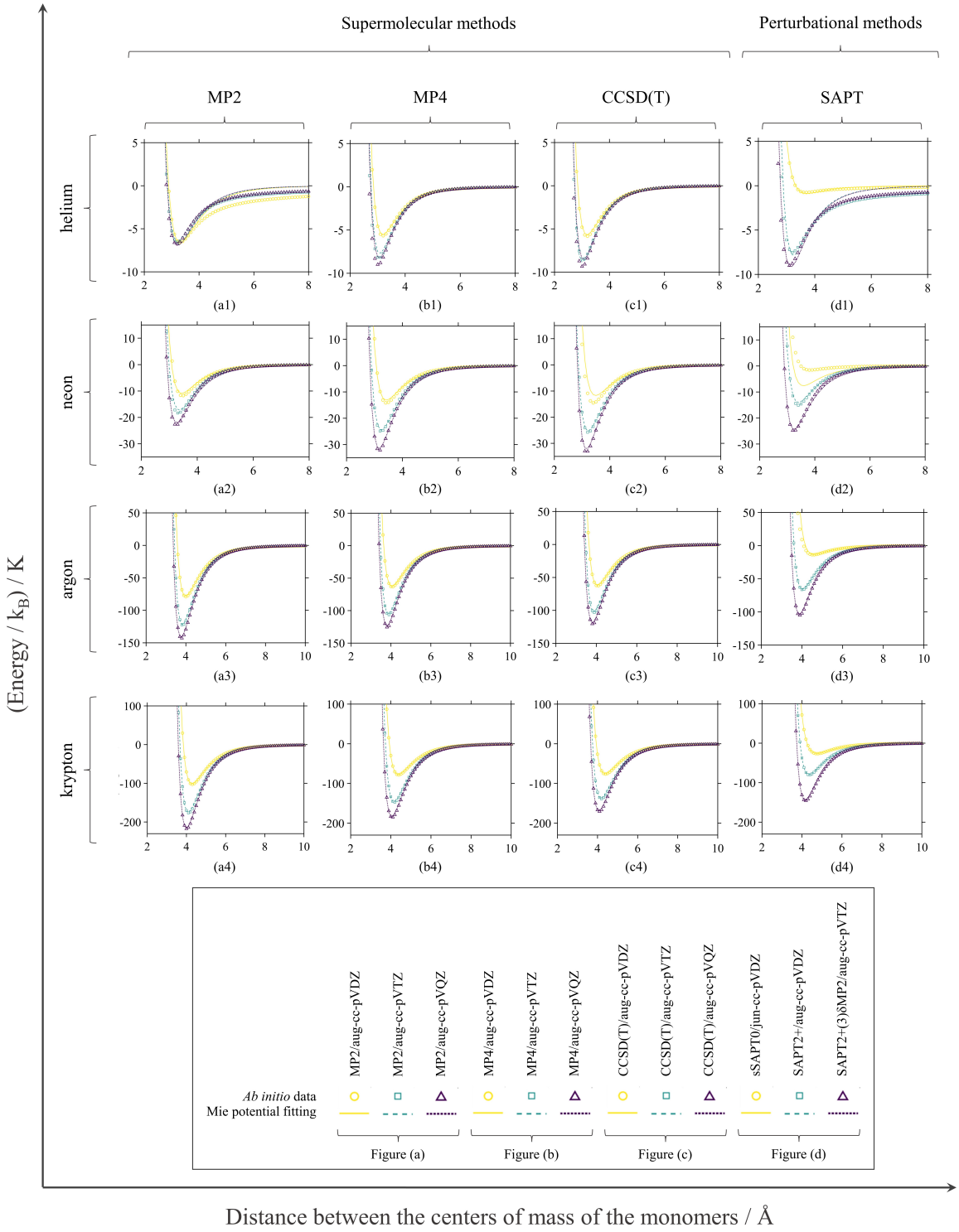

Figure 4 – Comparison of Mie potential energy curves obtained from bottom-up methodology parameters and *ab initio* interaction energy data obtained via supermolecular and perturbational methods for some noble gases. Second-order Møller-Plesset (MP2) energy results for (a1) helium, (a2) neon, (a3) argon, and (a4) krypton. Fourth-order Møller-Plesset (MP2) energy results for (b1) helium, (b2) neon, (b3) argon, and (b4) krypton. Coupled-cluster with singles, doubles, and perturbative triples excitations (CCSD(T)) energy results for (c1) helium, (c2) neon, (c3) argon, and (c4) krypton. Symmetry-adapted perturbation theory (SAPT) energy results for (d1) helium, (d2) neon, (d3) argon, and (d4) krypton.

Comparing the results obtained presented in Table 1 and Fig. 4, the depth of the potential energy well increases as the number of zeta basis functions in valence atomic orbitals increases for bottom-up-based Mie potential parameters in supermolecular methods. Increasing from double to triple-zeta basis functions (*i.e.*, aug-cc-pVDZ to aug-cc-pVTZ) and from triple to quadruple-zeta basis functions (*i.e.*, aug-cc-pVTZ to aug-cc-pVQZ) results in an increase in ε but a decrease in σ. The increase in ε is more significant when going from double to triple-zeta basis functions than from triple to quadruple-zeta basis functions.

Analyzing different post-Hartree-Fock methods considered in supermolecular methods (*i.e.*, MP2, MP4, and CCSD(T)), no pattern in the variation of bottom-up-based Mie potential parameters obtained was observed. Increasing the SAPT levels from bronze to silver (*i.e.* sSAPT0/jun-cc-pVDZ to SAPT2+/aug-cc-pVDZ) and from silver to gold (*i.e.*, SAPT2+/aug-cc-pVDZ to SAPT2+(3)δMP2 /aug-cc-pVTZ) standards results in an increase in the value of ε and a reduction in the σ value of the Mie potential.

The limited top-down and bottom-up-based Mie potential parameters found in the literature [10, 103, 104] are in agreement with the bottom-up-based ones obtained in this study by supermolecular methods considering triple as quadruple zeta basis functions and those obtained by perturbational methods for the silver and gold standards of SAPT. The top-down and bottom-up-based Mie potential parameters of Table 1 were used in the SAFT-VR Mie equation of state to calculate thermophysical properties considering NIST reference data listed in Table C1 found in Appendix C.

The SAFT-VR Mie equation of state results were compared to NIST reference data [95]. The accuracy for predicting NIST thermophysical properties data such as density (ρ), isochoric heat capacity (c_{v}), isobaric heat capacity (c_{p}), speed of sound (c_{sound}), and Joule-Thomson coefficient (μ_{JT}) was evaluated by computing AARD% values by Eq. (2.114) available in Section 2.2.2. The AARD% values obtained for the thermophysical properties considered for some noble gases using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state are presented in Fig. 5. Alternatively, the information presented in Fig. 5 is also available in Table D1 in Appendix D.

For all set of Mie potential parameters presented in Fig. 5, the lowest AARD% values were obtained for thermophysical properties ρ, c_{v}, c_{p}, and c_{sound}. The AARD% values for μ_{JT} obtained were high, in some cases reaching hundreds of percent. As expected, the AARD% values for the top-down-based Mie potential parameters were low since, in this methodology, the parameters are directly fitted to reproduce experimental data. Fig. 5 indicated that the bottom-up-based Mie potential parameters can also be used to predict thermophysical properties with certain accuracy, given the low AARD% values obtained.

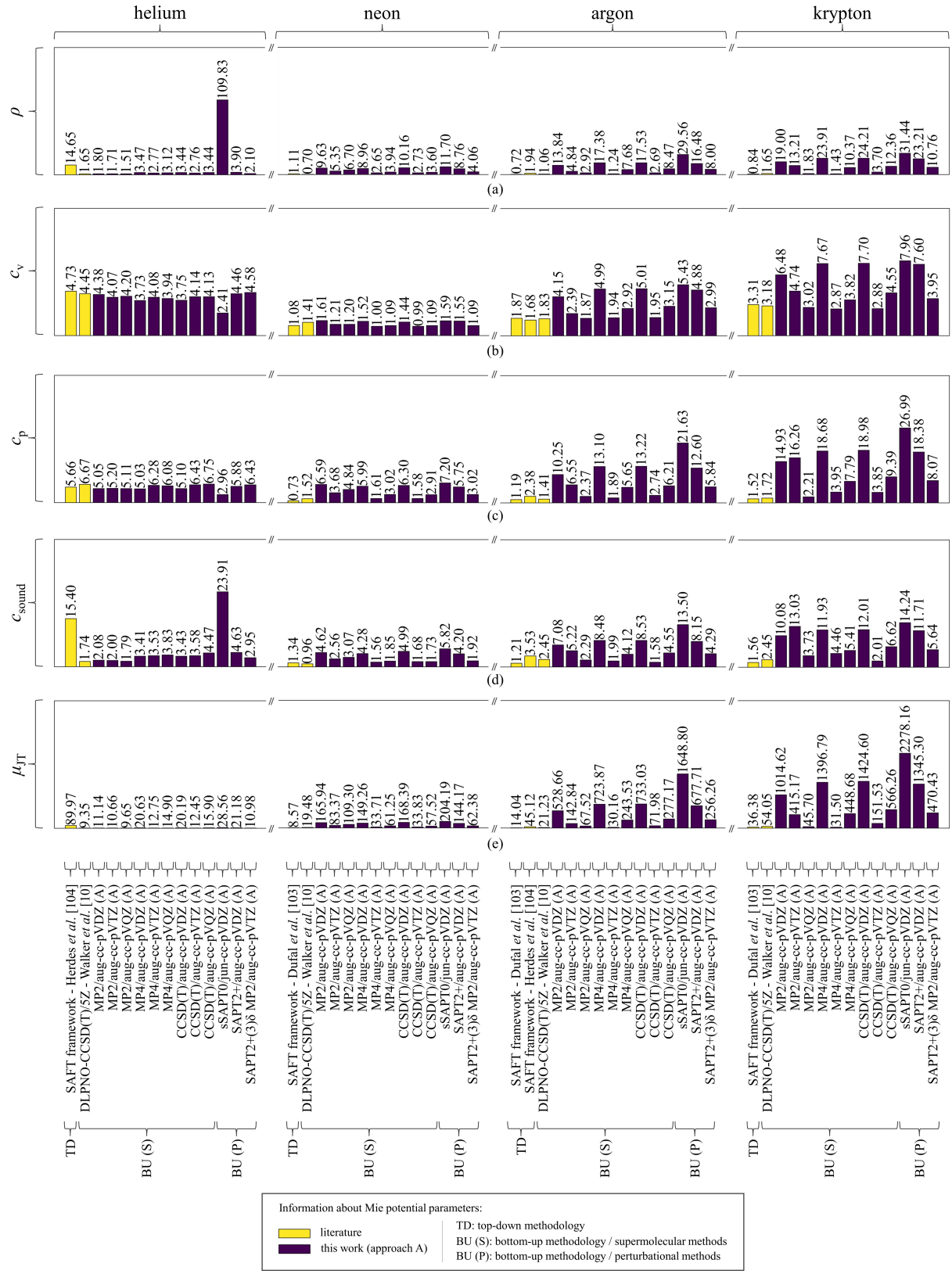

Figure 5 – AARD% values for thermophysical properties for some noble gases (helium, neon, argon, and krypton) using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state. Results for (a) ρ, (b) c_v, (c) c_p, (d) c_{sound}, and (e) μ_{JT}.

In supermolecular methods, the AARD% values reduce when increasing from double to triple or quadruple-zeta basis functions (*i.e.*, from aug-cc-pVDZ to aug- cc-pVTZ or aug-cc-pVQZ) in the bottom-up-based Mie potential parameters. There is no significant reduction in AARD% values when increasing from triple to quadruple-zeta basis functions (*i.e.*, from aug-cc-pVTZ to aug-cc-pVQZ). These observations suggest that the values of the Mie potential parameter ε are underestimated and the values of σ are overestimated when using double-zeta basis functions (*i.e.*, aug-cc-pVDZ), given the higher AARD% values obtained.

When using the same basis function, the change in the post-Hartree-Fock methods considered in supermolecular methods (*i.e.*, MP2, MP4, and CCSD(T)) does not result in very significant changes in AARD% values. In perturbational methods, a notable decrease is observed in AARD% values as the SAPT level increases from bronze to silver (*i.e.* sSAPT0/jun-cc-pVDZ to SAPT2+/aug-cc-pVDZ) and from silver to gold (*i.e.*, SAPT2+/aug-cc-pVDZ to SAPT2+(3)δMP2 /aug-cc-pVTZ) standards. The AARD% values obtained for the gold standard of SAPT are comparable to those obtained through supermolecular methods using triple-zeta basis functions (*i.e.*, from aug-cc-pVTZ).

The top-down and bottom-up-based Mie potential parameters presented in Table were also used in the SAFT-VR Mie equation of state to calculate the critical point of some noble gases considered in this study. Fig. 6 presents the percentage deviation for critical temperature (T_{crit}), pressure (p_{crit}), and density (ρ_{crit}) calculated relative to NIST reference data [95]. Alternatively, the information presented in Fig. 6 is also available in Table D2 in Appendix D. The results presented in Fig. 6 confirm that the trend observed in the AARD% values for thermophysical properties is also verified in the critical point. The top-down-based Mie potential parameters result in lower percentage deviations from NIST reference data since these parameters are directly fitted in the SAFT-VR Mie equation of state to reproduce experimental data. Excluding helium, bottom-up-based Mie potential parameters showed percentage deviations in the tens of percent and followed the same trends of results observed for the AARD% values considering different methodologies analyzed.

In this study, the MP2/aug-cc-pVTZ Mie potential parameters were found to produce better predictions for thermophysical properties and critical point with lower computational cost. Thus, the set Mie potential parameters from this theory level were chosen to represent the prediction of the thermophysical properties of some noble gases using the SAFT-VR Mie equation of state. The prediction results are presented in Fig. 7. According to Fig. 7, the chosen set of Mie potential parameters can reproduce NIST reference data with certain accuracy. Fig. 7 demonstrates that the Mie potential parameters derived from *ab initio* calculations, without the need for experimental data, can be applied in a state equation such as the SAFT-VR Mie and still provide reliable predictions.

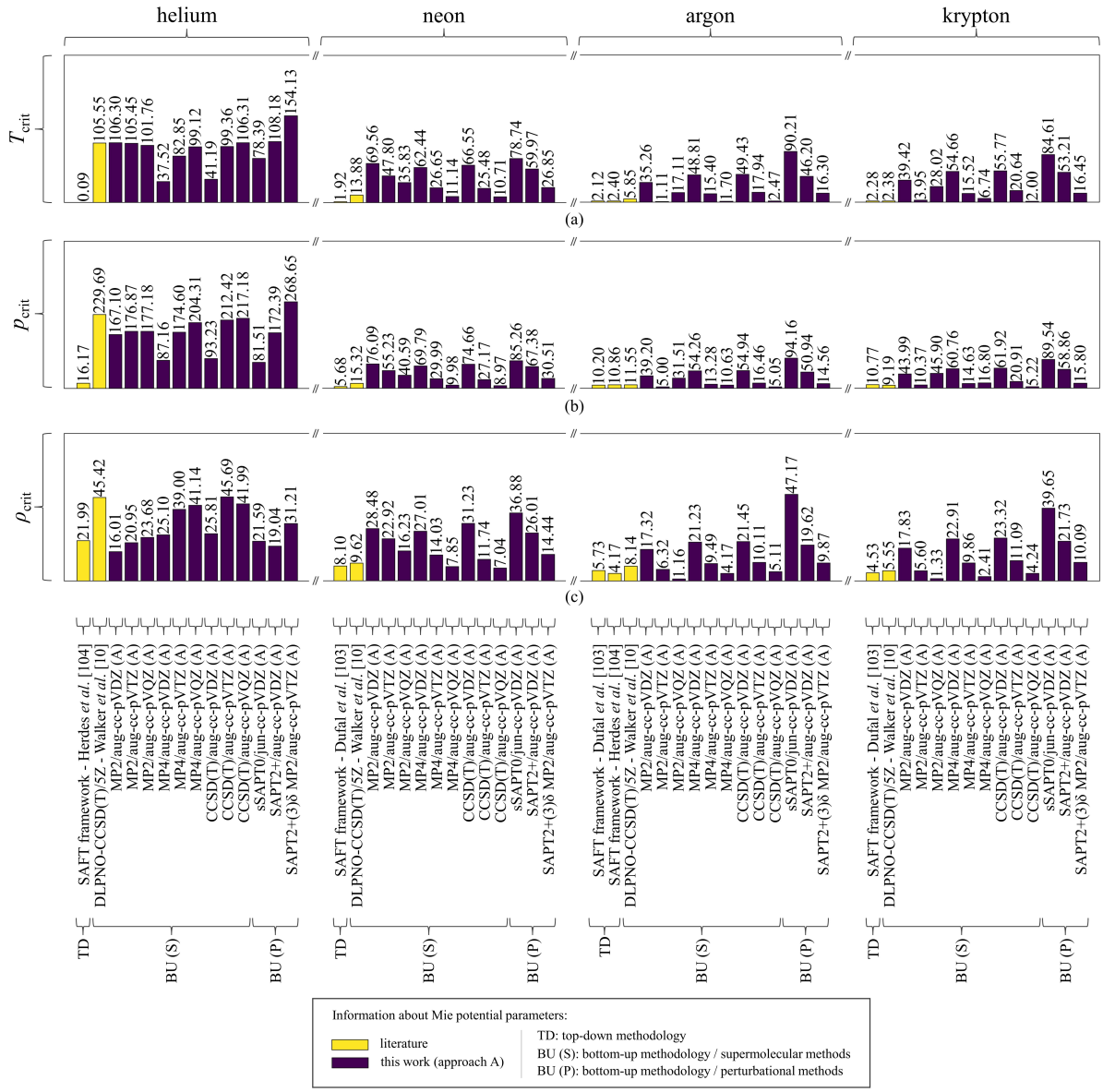

Figure 6 – Percentage deviation of the calculated critical point from NIST reference data using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state for some noble gases (helium, neon, argon, and krypton). Results for (a) T_{crit}, (b) p_{crit}, and (c) ρ_{crit}.

The bottom-up-based Mie potential parameters of MP2/aug-cc-pVTZ theory level were compared with those bottom-up and top-down-based ones available in the literature for the prediction of saturation properties such as saturation pressure (p_{sat}) and enthalpy of vaporization (ΔH_{vap}). Fig. 8 shows the AARD% values obtained for the saturation properties when considering argon and krypton. Alternatively, the information presented in Fig. 8 is also available in Table D3 in Appendix D. The AARD% values presented in Fig. 8 are computed considering NIST reference data [95] listed in Table C2 found in Appendix C. In Fig. 8, only argon and krypton are presented since for helium the critical temperature value is close to absolute zero (which caused problems in the calculations) and for neon there is a very few number of NIST reference data available up to the computed critical temperature for MP2/aug-cc-pVTZ Mie potential

parameters.

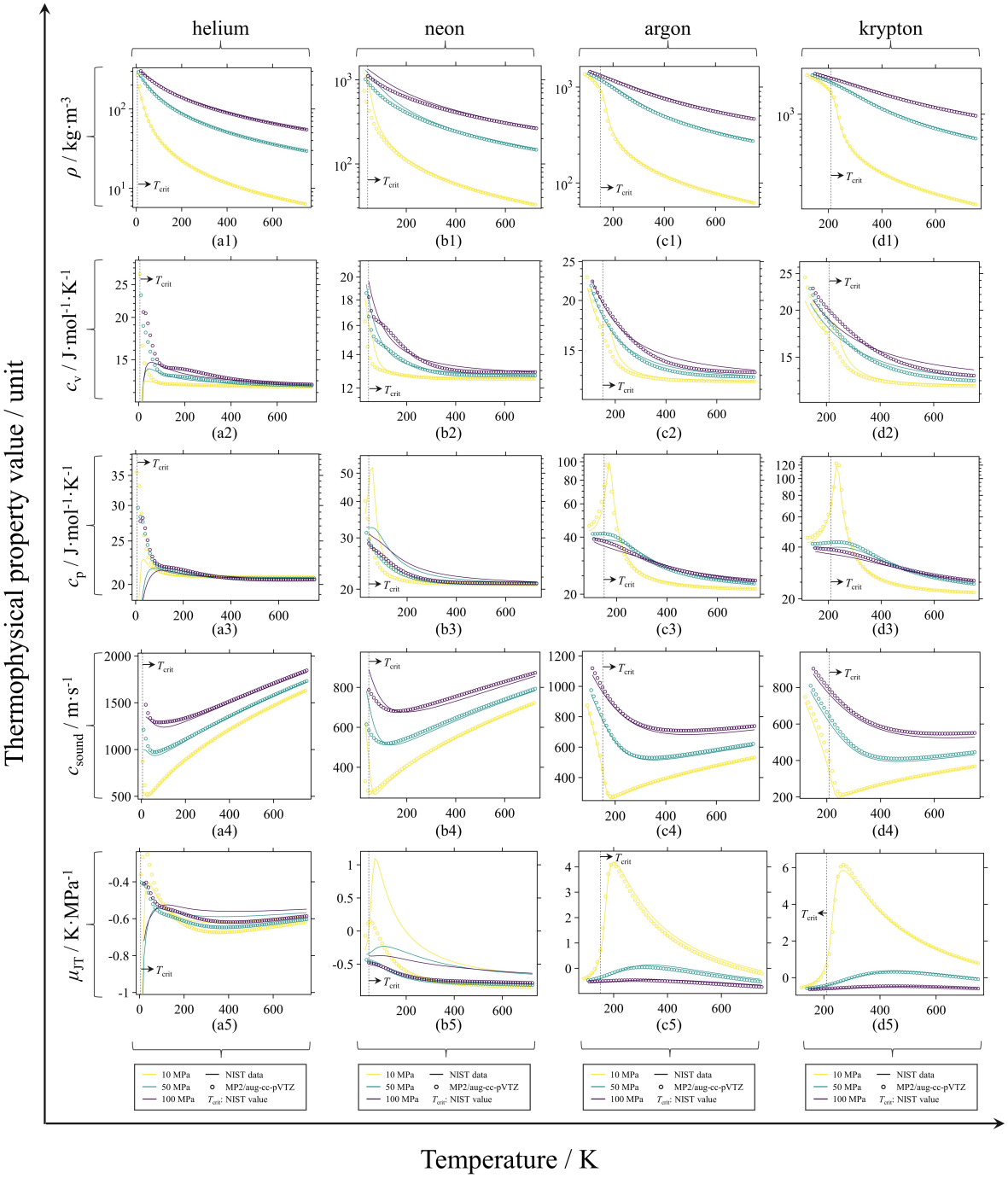

Figure 7 – Predicted thermophysical properties for some noble gases using bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies by approach A in the SAFT-VR Mie equation of state. Helium results for (a1) ρ, (a2) c_v, (a3) c_p, (a4) c_{sound}, and (a5) μ_{JT}. Neon results for (b1) ρ, (b2) c_v, (b3) c_p, (b4) c_{sound}, and (b5) μ_{JT}. Argon results for (c1) ρ, (c2) c_v, (c3) c_p, (c4) c_{sound}, and (c5) μ_{JT}. Krypton results for (d1) ρ, (d2) c_v, (d3) c_p, (d4) c_{sound}, and (d5) μ_{JT}.

Fig. 7, which refers to the prediction of thermophysical properties, indicates that with the reduction in temperature towards the critical temperature, the parameters of the bottom-up methodology lose some of their predictive capacity. As saturation property data are obtained at

lower temperatures (below the critical temperature), deviations from NIST reference data are expected, confirmed through Fig. 8. Even so, the AARD% values presented in Fig. 8 are in the order of tens of percent for a set of parameters obtained through the bottom-up methodology, without considering a fitting to data experimental. To complement the results of Fig. 8, Fig. 9 illustrates the predictive capacity of saturation properties using the Mie potential parameters obtained by the bottom-up methodology through the MP2/aug-cc-pVTZ theory level.

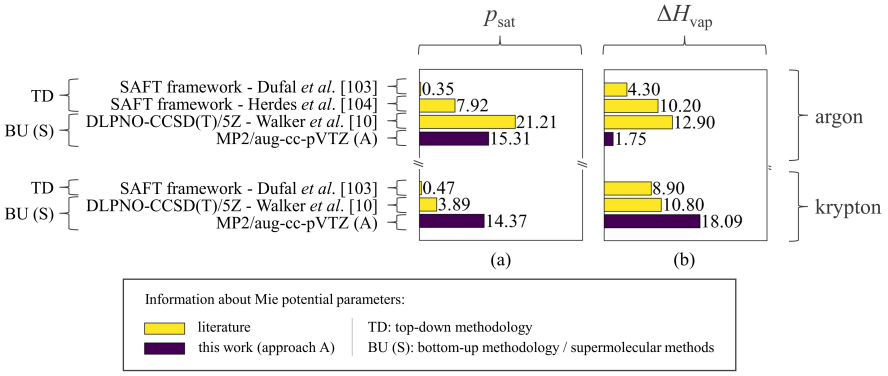

Figure 8 – AARD% values for saturation properties for some noble gases using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state. Results of (a) p_{sat} and (b) ΔH_{vap} for argon and krypton.

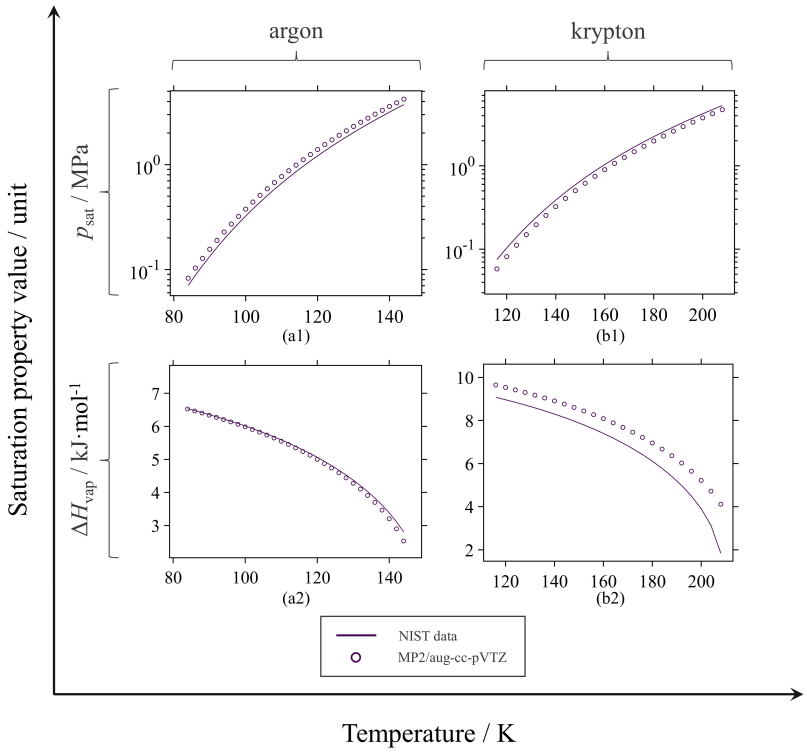

Figure 9 – Predicted saturation properties for some noble gases using bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies by approach A in the SAFT-VR Mie equation of state. Results of p_{sat} for (a1) argon and (b1) krypton. Results of ΔH_{vap} for (a2) argon and (b2) krypton.

3.2 Second group: methane and substituted-methane compounds

Some of the results presented in the subsection of this study have already been published. If you use the content presented here, kindly cite the reference [46]:

LYRA, E.P.; FRANCO, L. F. M. Deriving force fields with a multiscale approach: From ab initio calculations to molecular-based equations of state. The Journal of Chemical Physics, v. 157, n. 11, p. 114107, 2002. doi: 10.1063/5.0109350

3.2.1 Methane modeling

First, the geometry of the methane molecule was optimized. For the MP2/aug-cc-pVQZ theory level, a tetrahedral geometry was found with a carbon-hydrogen bond length equal to 1.0845 Å, which is close to the value of 1.087±0.001 Å reported by Hirota [105]. The optimized methane molecule was used to construct 12 symmetric conformers considered in the work of Chao *et al*. [92], as previously described in Fig. 3(b) presented in Section 2.2.2 of this study or in the original publication.

Supermolecular and perturbational methods were used to compute the interaction energy curves for the 12 symmetric methane conformers. Applying supermolecular methods, the interaction energies were obtained through second and fourth-order Møller-Plesset (*i.e.*, MP2 and MP4 energies, respectively) and coupled-cluster with singles, doubles, and perturbative triples excitations (*i.e.,* CCSD(T) energy). MP2 energies were obtained using Dunning basis sets for double, triple, and quadruple-zeta (*i.e.*, aug-cc-pVnZ, with n = D, T, and Q, respectively). Due to high computational costs, MP4 and CCSD(T) energies were obtained using Dunning basis sets for double-zeta (*i.e.*, aug-cc-pVDZ). basis sets for double-zeta (*i.e.*, aug-cc-pVnZ, with n = D). By perturbational methods, the interaction energies were obtained through bronze, silver, and gold standards of SAPT (*i.e.*, sSAPT0/jun-cc-pVDZ, SAPT2+/aug-cc-pVDZ, and SAPT2+(3)δMP2/aug-cc-pVTZ energies, respectively).

The methane molecule was represented by a spherical 1-site coarse-grained model based on the Mie potential. The Mie potential parameters σ, ε, and λ^{rep} were fitted to represent interaction energy curves for the 12 symmetric conformers obtained by supermolecular and perturbational methods. A single set of Mie potential parameters for methane was obtained considering the two approaches, A and B. To obtain the Mie potential parameters, the objective functions to be minimized for approaches A and B were Eqs. (2.110) and (2.112), respectively, presented in Section 2.2.2 of this study.

Top-down and bottom-up-based Mie potential parameters considered in this study are presented in Table 2. The top-down-based Mie potential parameters available in the literature are those from the SAFT-γ Mie force field [106, 107] and those optimized for the SAFT-VR Mie equation of state by Dufal *et al.* [103].

Table 2 presents the Mie potential parameters obtained for methane through the top-down and bottom-up-based Mie potential parameters. The top-down-based Mie potential parameters available in the literature are those from the SAFT framework (*i.e.*, SAFT-γ Mie force field and SAFT-VR Mie equation of state). For methane, only the set of Mie potential parameters obtained by Walker *et al.* [10] are from the bottom-up methodology within the SAFT framework.

Table 2 – Top-down-based Mie potential parameters within the SAFT framework available in the literature and bottom-up-based Mie potential parameters available in the literature and those obtained in this study from *ab initio* data by different methodologies and contexts for methane.

		Mie potential parameters[a,b,f]					
Methodology[c]	Context[d]	m	σ/Å	λ_{rep}	λ_{att}	(ε/k_B)/K	Reference
TD	SAFT framework	1.0000	3.7525	16.4191	6.0000	170.9317	[e]
	SAFT framework	1.0000	3.7366	12.3190	6.0000	151.4699	[103]
BU (S)	DLPNO-CCSD(T)/5Z	1.0000	3.8495	12.8260	6.0000	140.9100	[10]
	MP2/aug-cc-pVDZ (A)	1.0000	3.8114	12.1176	6.0000	116.7706	this work
	MP2/aug-cc-pVDZ (B)	1.0000	3.4182	12.0000	6.0000	142.2670	this work
	MP2/aug-cc-pVTZ (A)	1.0000	3.7236	12.1942	6.0000	145.5861	this work
	MP2/aug-cc-pVTZ (B)	1.0000	3.3362	12.0000	6.0000	177.8212	this work
	MP2/aug-cc-pVQZ (A)	1.0000	3.7042	12.2061	6.0000	153.2803	this work
	MP2/aug-cc-pVQZ (B)	1.0000	3.3147	12.0000	6.0000	188.4825	this work
	MP4/aug-cc-pVDZ (A)	1.0000	3.7683	12.1480	6.0000	135.0757	this work
	MP4/aug-cc-pVDZ (B)	1.0000	3.3895	12.0000	6.0000	164.5450	this work
	CCSD(T)/aug-cc-pVDZ (A)	1.0000	3.7751	12.1445	6.0000	132.5110	this work
	CCSD(T)/aug-cc-pVDZ (B)	1.0000	3.3939	12.0000	6.0000	161.4271	this work
BU (P)	sSAPT0/jun-cc-pVDZ (A)	1.0000	3.9868	12.1253	6.0000	58.6368	this work
	sSAPT0/jun-cc-pVDZ (B)	1.0000	3.6330	12.0000	6.0000	57.2287	this work
	SAPT2+/aug-cc-pVDZ (A)	1.0000	3.7182	12.1521	6.0000	152.4757	this work
	SAPT2+/aug-cc-pVDZ (B)	1.0000	3.3287	12.0000	6.0000	189.1362	this work
	SAPT2+(3)δMP2/aug-cc-pVTZ (A)	1.0000	3.6964	12.2202	6.0000	161.4271	this work
	SAPT2+(3)δMP2/aug-cc-pVTZ (B)	1.0000	3.3273	12.0000	6.0000	195.8246	this work

[a]Methane is represented as spherical 1-site coarse-grained model (in the SAFT-VR Mie equation of state is equivalent to assigning the number of segments m equal to 1). [b]k_B is the Boltzmann constant. [c]TD and BU refer to the top-down and bottom-up methodologies, respectively, while (S) and (P) refer to the supermolecular and perturbational methods, respectively. [d](A) and (B) indicate that approaches A and B objective functions, respectively, were minimized in the parameter fitting procedure. [e]Professor Erich A. Müller (personal communication, June 3rd, 2021), based on published works [106, 107]. [f]The best results for fitting the parameters of approach (B) were observed by setting the λ_{rep} value equal to 12.

The results in Table 2 show that approach A for bottom-up methodology yields values for the Mie potential parameters ε and σ closer to those reported for the top-down methodology found in the literature within the SAFT framework. The Mie potential parameter values obtained by approach A through the bottom-up methodology in this study are also consistent with those available in the literature for the bottom-up methodology. According to Table 2, approach B leads to higher ε values and lower σ values than approach A. This suggests that approach B gives more weight to more stable conformers during the parameter fitting procedure. In more stable conformers, steric hindrances are smaller, which allows the dimer molecules to come closer together.

In Fig. 10, interaction energy curves obtained by *ab initio* calculations and Mie potential energy curves obtained using parameters from top-down methodology and those from the bottom-up methodology arising from this study are presented. Fig. 10 shows a very interesting trend. The Mie potential energy curves obtained using parameters from the top-down methodology are close to the average of the *ab initio* data used in acquiring the parameters from the bottom-up methodology. The gray shaded areas in Fig. 10 highlight this trend, which represent the coverage area of data obtained using supermolecular and perturbational methods. This observation indicates that bottom-up-based Mie potential parameters obtained by approach A result in values of σ and ε closer to those verified by the top-down methodology within the SAFT framework.

Fig. 11 demonstrates the interaction energy curves by *ab initio* calculations for each of the 12 symmetric conformers studied by Chao *et al.* [92]. The interaction energy curves for each conformer were found to be distinct, indicating that not all 12 conformers are equally accessible. This observation implies that considering anisotropic effects in the parameter fitting procedure as in approaches A and B becomes justifiable. A more detailed discussion of the differences between these interaction energy curves is provided by Chao *et al.* [92].

The thermophysical properties of methane were computed using top-down and bottom-up-based Mie potential parameters. This was done through the use of the SAFT-VR Mie equation of state and by molecular dynamics simulations. Density (ρ), isochoric heat capacity (c_{v}), isobaric heat capacity (c_{p}), and speed of sound (c_{sound}) were computed for a dataset comprising a series of temperature and pressure conditions. The computed thermophysical properties were compared with NIST reference data. The AARD% value was used as a metric to evaluate the accuracy of predicting the results of thermophysical properties. Fig. 12 shows the AARD% values for methane obtained for each thermophysical property evaluated using top-down and bottom-up-based Mie potential parameters. The AARD% values presented in Fig. 12 were obtained considering NIST reference data listed in Table C1 found in Appendix C. Alternatively, the information presented in Fig. 12 is also available in Table D4 in Appendix D.

According to Fig. 12, using top-down-based Mie potential parameters results in lower AARD% values when reproducing thermophysical properties such as ρ, c_v, c_p, and c_{sound}. This observation is expected since in the top-down methodology the Mie potential parameters are directly fitted to represent experimental data. As a result of the low AARD% values obtained using top-down-based Mie potential parameters, the spherical 1-site coarse-grained model is a suitable representation for the methane molecule.

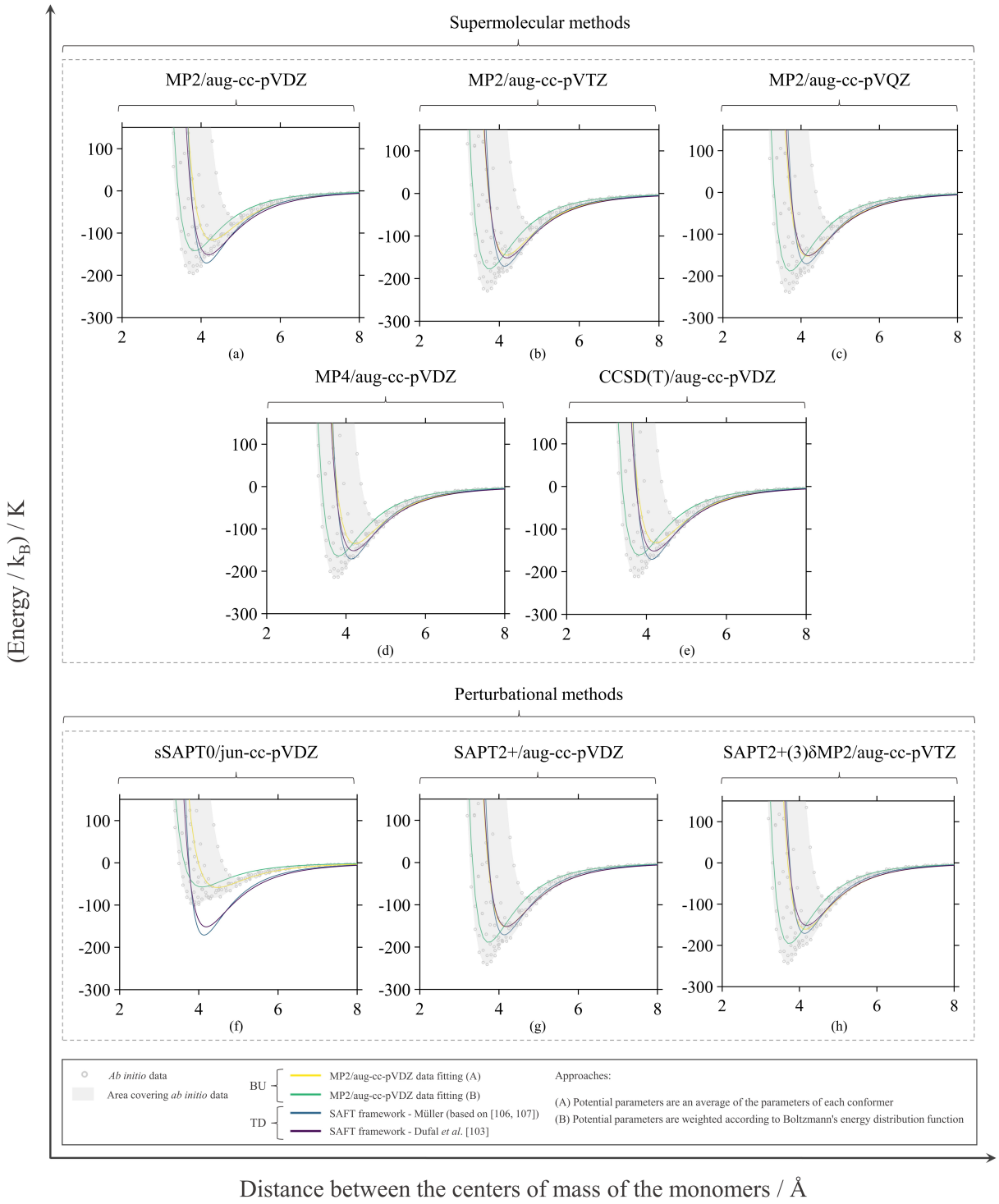

Figure 10 – Comparison of Mie potential energy curves obtained using parameters from top-down (TD) and bottom-up (BU) methodologies and *ab initio* interaction energy data obtained via supermolecular and perturbational methods for methane. Results for supermolecular methods considering (a) MP2/aug-cc-pVDZ, (b) MP2/aug-cc-pVTZ, (c) MP2/aug-cc-pVQZ, (d) MP4/aug-cc-pVDZ, and (e) CCSD(T)/aug-cc-pVDZ energies. Results for perturbational methods considering (f) sSAPT0/jun-cc-pVDZ, (g) SAPT2+/aug-cc-pVDZ, and (h) SAPT2+(3)δMP2/aug-cc-pVTZ energies.

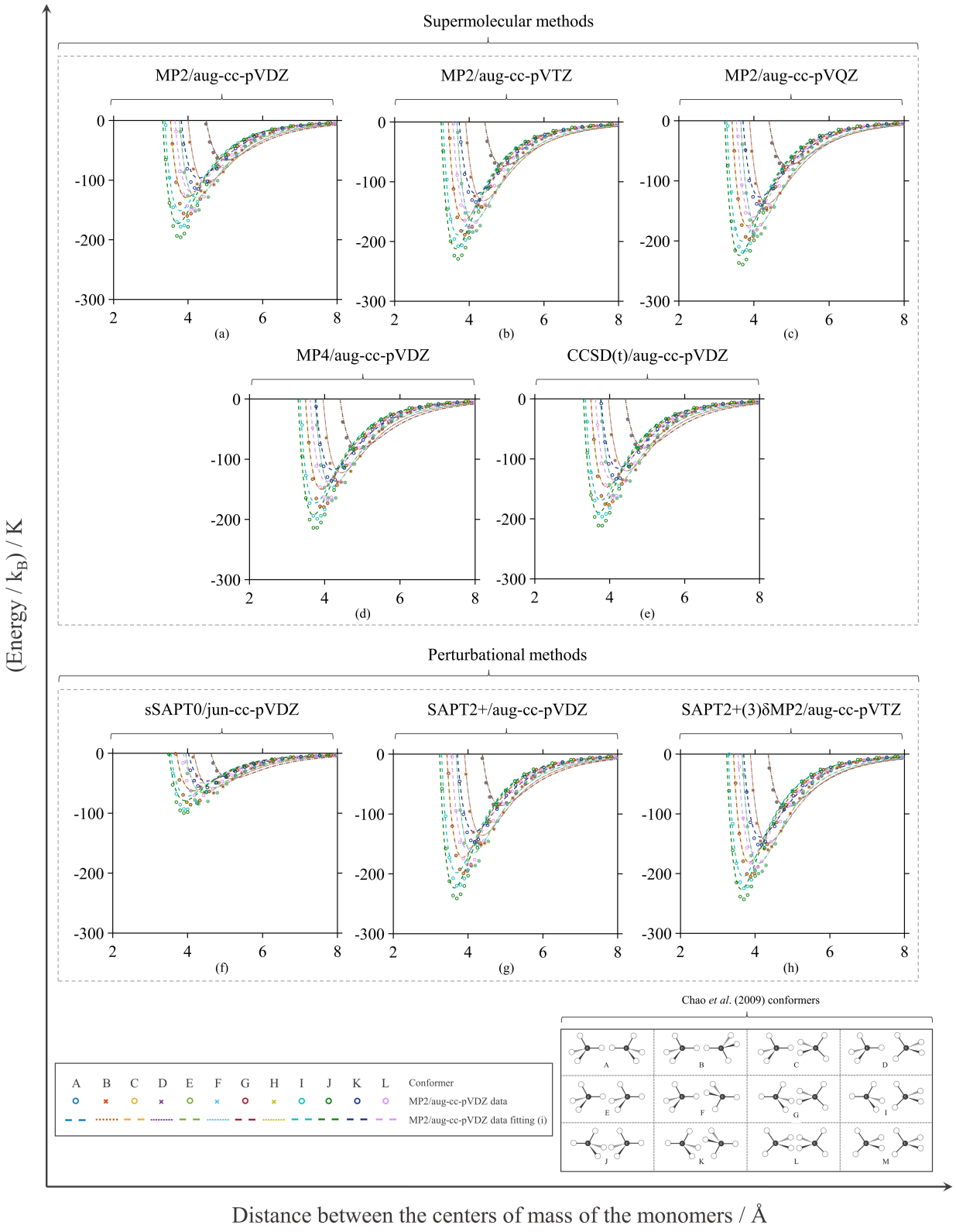

Figure 11 – The Mie potential energy curves for approach A and *ab initio* interaction energy data obtained via supermolecular and perturbational methods for each methane conformer. Results for supermolecular methods considering (a) MP2/aug-cc-pVDZ, (b) MP2/aug-cc-pVTZ, (c) MP2/aug-cc-pVQZ, (d) MP4/aug-cc-pVDZ, and (e) CCSD(T)/aug-cc-pVDZ energies. Results for perturbational methods considering (f) sSAPT0/jun-cc-pVDZ, (g) SAPT2+/aug-cc-pVDZ, and (h) SAPT2+(3)δMP2/aug-cc-pVTZ energies.

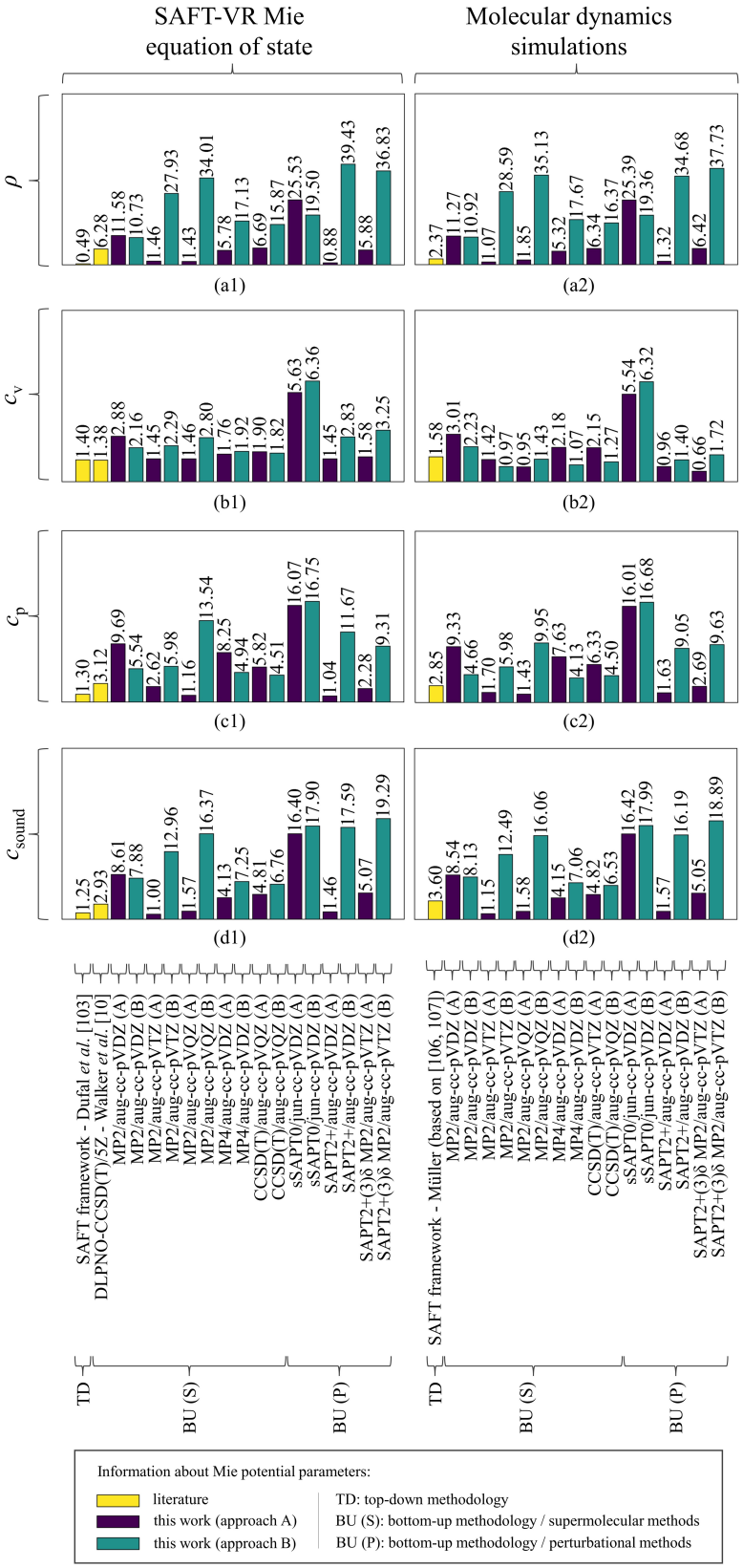

Figure 12 – AARD% values for thermophysical properties for methane using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state and molecular dynamics simulations. Results of (a1) ρ, (b1) c_v, (c1) c_p, and (d1) c_{sound} by SAFT-VR Mie equation of state. Results of (a2) ρ, (b2) c_v, (c2) c_p, and (d2) c_{sound} by molecular dynamics simulations.

The AARD% values presented in Fig. 12 also demonstrate that bottom-up-based Mie potential parameters for approach A can be used to predict with a certain accuracy thermophysical properties of methane using the SAFT-VR Mie equation of state and molecular dynamics simulations. The lowest values of AARD% obtained by supermolecular methods were obtained using the MP2/aug-cc-pVTZ and MP2/aug-cc-pVQZ theory levels, in which the increase in the number of zeta basis functions did not significantly improve accuracy. The increase in the level of theory in supermolecular methods from MP2 to MP4 or CCSD(T) for the same basis set (*i.e.*, aug-cc-pVDZ) reduced AARD% values with an increase in computational cost. Comparing the AARD% values obtained through bottom-up methodology by perturbational methods, the lowest values were obtained for SAPT2+/aug-cc-pVDZ and SAPT2+(3)δMP2/aug-cc- pVTZ levels. Increasing the SAPT level in perturbational methods from SAPT2+/aug-cc-pVDZ to SAPT2+(3)δMP2/aug-cc-pVTZ did not significantly reduce the AARD% values.

According to Fig. 12, the AARD% values obtained from using the SAFT-VR Mie equation of state and molecular dynamics simulations are consistent. Both computational methods were tested for computing thermophysical properties and presented very similar accuracy results. The SAFT-VR Mie equation of state is a model with a series of approximations that simplify the description of fluid structure, whereas molecular dynamics simulations were expected to produce lower AARD% values when predicting thermophysical properties.

When comparing the AARD% values obtained for approaches A and B of the bottom-up methodology, approach A demonstrated to yield lower AARD% values for reproducing thermophysical properties. This observation was expected since the Mie potential parameter values obtained by approach A were generally closer to the top-down methodology ones. This trend is observed with the exception of MP2/aug-cc-pVDZ and sSAPT0/jun-cc-pVDZ results.

Critical temperature (T_{crit}), pressure (p_{crit}), and density (ρ_{crit}) of methane were computed using the top-down and bottom-up-based Mie potential parameters. The critical point calculation was performed using the SAFT-VR Mie equation of state. In Fig. 13, methane critical points are compared to NIST reference data through the percentage deviation. Alternatively, the information presented in Fig. 13 is also available in Table D5 in Appendix D. The results of Fig. 13 indicate that using top-down-based Mie potential parameters lower percentage deviations to the NIST reference data. Within the bottom-up methodology, approach A is better at reproducing the critical point than approach B, as it presents lower percentage deviation values to the NIST reference data. The trends of the results observed for the critical point are in agreement with those observed for thermophysical properties.

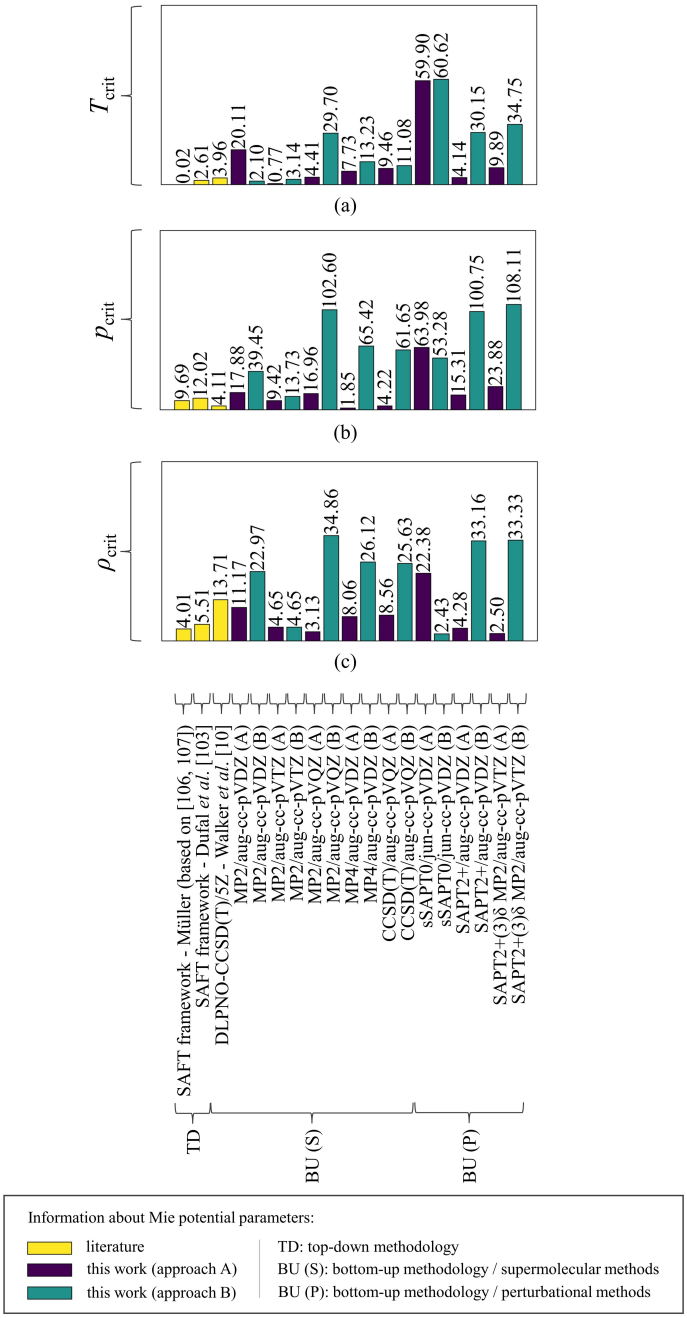

Figure 13 – Percentage deviation of the calculated critical point from NIST reference data using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state for methane. Results of (a) T_{crit}, (b) p_{crit}, and (c) ρ_{crit} for methane.

From the results previously presented for the bottom-up methodology, the sets of Mie potential parameters that best reproduced NIST reference data with lower computational cost for thermophysical properties and methane critical points were those obtained for MP2/aug-cc-pVTZ and SAPT2+/aug-cc-pVDZ using approach A. These sets of parameters presented lower values of AARD% for thermophysical properties and lower percentage deviations for critical point. To exemplify the predictive capacity of parameters obtained by the bottom-up approach, the Mie potential parameters for the MP2/aug-cc-pVTZ theory level obtained by approach A were chosen to predict thermophysical property data for methane.

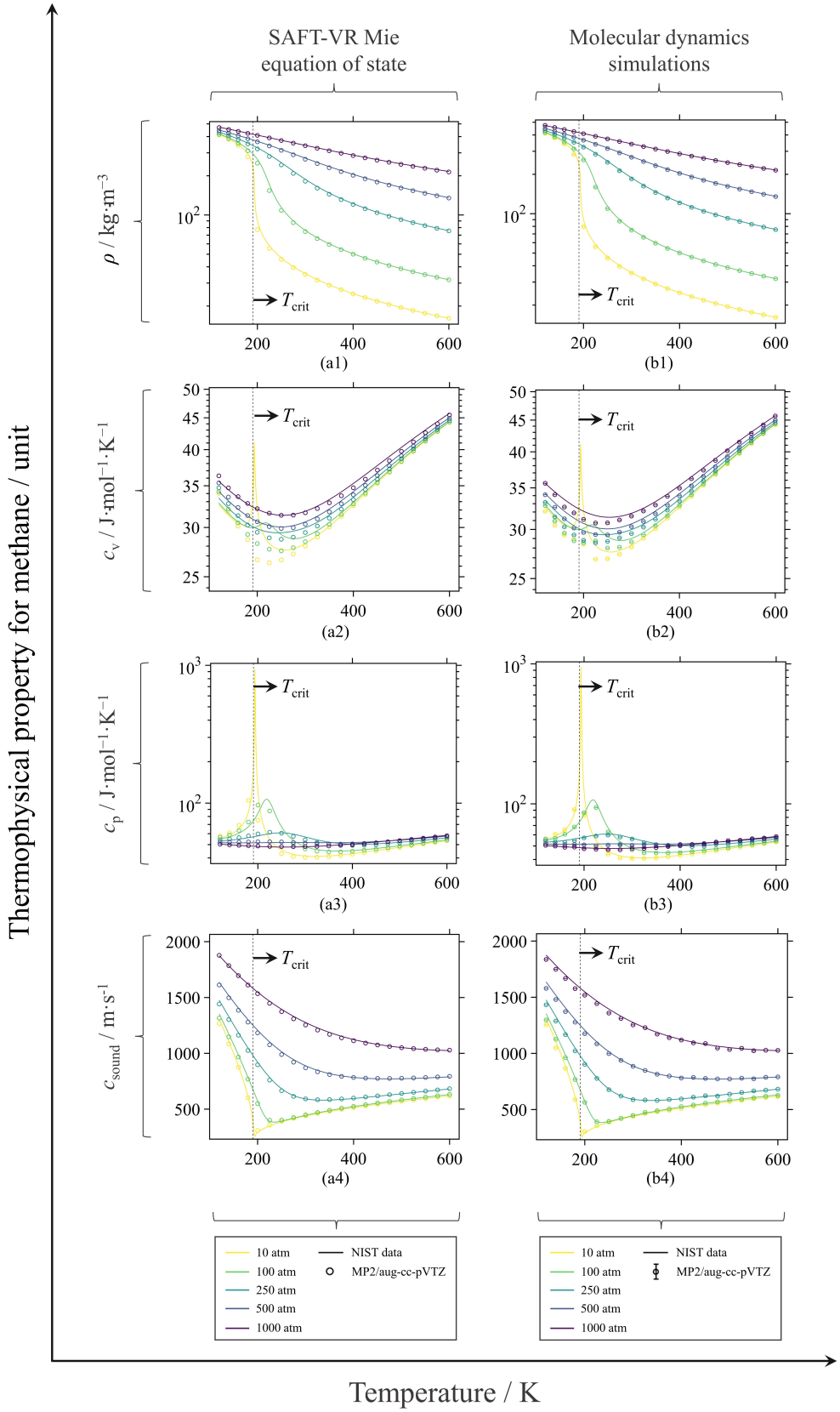

Figure 14 – Predicted thermophysical properties for methane using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Prediction of (a1) ρ, (b1) c_v, (c1) c_p, and (d1) c_{sound} by SAFT-VR Mie equation of state. Prediction of (a2) ρ, (b2) c_v, (c2) c_p, and (d2) c_{sound} by molecular dynamics simulations.

The predictive results for thermophysical properties of methane considering a spherical 1-site coarse-grained model using the bottom-up-based Mie potential parameters obtained through MP2/aug-cc-pVTZ theory level energies by approach A is presented in Fig. 14. The results obtained via SAFT-VR Mie equation of state and molecular dynamics simulations presented in Figs. 14(a) and 14(b) are compared with NIST reference data. The results presented in Fig. 14 demonstrate that the spherical 1-site coarse-grained model with Mie potential parameters derived from interaction energy curves obtained at MP2/aug-cc-pVTZ theory level by approach A reproduces NIST reference data very well. This observation is supported by the analysis performed for the AARD% values.

Saturation pressure (p_{sat}) and enthalpy of vaporization (ΔH_{vap}) were also computed using some bottom-up and top-down-based Mie potential parameters. The AARD% values obtained for the saturation properties calculated through SAFT-VR Mie equation of state for methane are presented in Fig. 15. Alternatively, the information presented in Fig. 15 is also available in Table D6 in Appendix D. The AARD% values presented in Fig. 15 were calculated considering NIST reference data listed in Table C2 found in Appendix C. Fig. 15 demonstrates that the AARD% values obtained for Mie potential parameters from the bottom-up methodology are considerably higher than for the top-down methodology, except for the values for ΔH_{vap} using MP2/aug-cc-pVTZ approach A Mie potential parameters.

In order to complement the results in Fig. 15, the predictive capacity of saturation properties for Mie potential parameters obtained by the bottom-up methodology through the MP2/aug-cc-pVTZ approach A is presented in Fig. 16. From Fig. 16, it can be seen a good adequacy of the Mie potential parameters obtained by the bottom-up methodology in this work when used in the SAFT-VR Mie equation of state to predict NIST saturation properties data. In addition to the fact that the Mie potential parameters of the top-down methodology are fitted to reproduce experimental data, it is noted that with the reduction of temperature values, the values of p_{sat} become very low, significantly impacting the calculation of AARD%. Therefore, it is essential to analyze AARD% values jointly with predicted property curves in comparison to reference data.

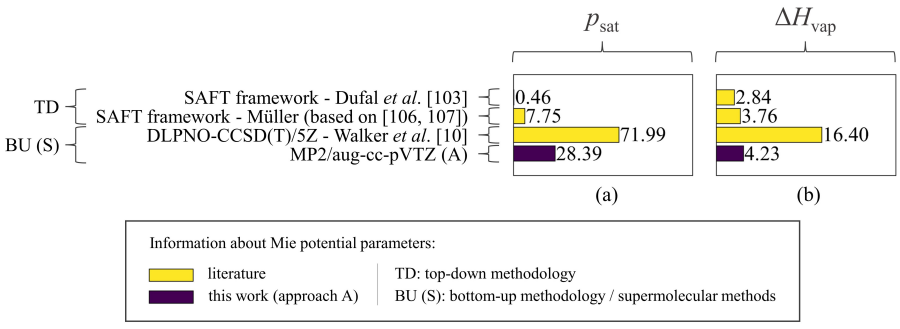

Figure 15 – AARD% values for saturation properties for methane using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state. Results of (a) p_{sat} and (b) ΔH_{vap} for methane.

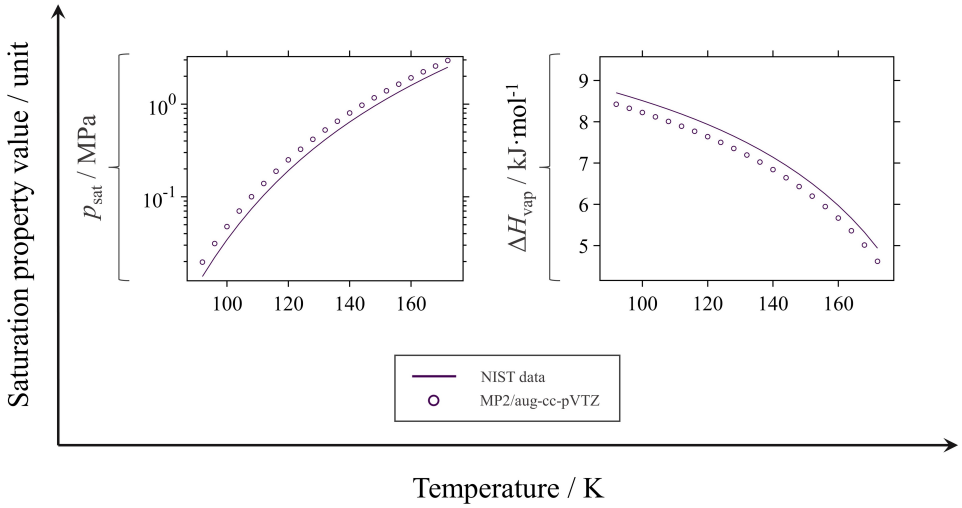

Figure 16 – Predicted saturation properties for methane using bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies by approach A in the SAFT-VR Mie equation of state. Methane results for (a) p_{sat} and (b) ΔH_{vap}.

3.2.2 Substituted-methane compounds modeling

A similar procedure to methane was adopted for some substituted-methane compounds (*i.e.*, refrigerants). Isolated molecules of tetrafluoromethane, fluoromethane, chlorotrifluoromethane, dichlorodifluoromethane, dichlorofluoromethane, and chlorodifluoromethane were also optimized at the MP2/aug-cc-pVQZ theory level. Of the molecules studied, only tetrafluoromethane presented tetrahedral geometry. The resulting equilibrium geometries are available in the publication derived from this dissertation available in the literature [46].

Isolated molecules of substituted-methane compounds with equilibrium geometry were used to build dimers based on the same 12 symmetric conformers considered for methane in the work of Chao *et al.* [92]. For molecules with different types of atoms bonded to carbon, such as fluoromethane, chlorotrifluoromethane, dichlorodifluoromethane, dichlorofluoromethane, and chlorodifluoromethane, these atoms bonded to carbon were also exchanged to generate new conformers. Additional information on the generated conformers can be found in Fig. 3(b) presented in Section 2.2.2 of this study or in the original publication

Due to the results obtained for methane and the computational cost involved in the calculations, interaction energy curves were built for some substituted-methane compounds by supermolecular methods using the MP2/aug-cc-pVTZ theory level data. A spherical 1-site coarse-grained model based on the Mie potential was also considered to represent substituted-methane compounds. The Mie potential parameters such as σ, ε, and λ^{rep} were obtained by minimizing the objective functions for approaches A and B given by Eqs. (2.110) and (2.112), respectively, presented in Section 2.2.2 of this study.

As some substituted-methane compounds of this study have atoms more electronegative than hydrogen bonded to the carbon atom (*i.e.*, chlorine and fluorine), the distortion of the electron cloud in these molecules is greater than in methane, which makes the steric effects also greater. Thus, approaches A and B are indirect ways of inserting a spatial dependence in the Mie potential parameters obtained through the bottom-up methodology.

Table 3 presents the top-down and bottom-up-based Mie potential parameters for some substituted-methane compounds of this study. According to Table 3, similarly to methane, approach A of bottom-up methodology also leads to the Mie potential parameters ε and σ closer to those from top-down methodology available in the literature within the SAFT framework. For some substituted-methane compounds studied, approach B of bottom-up methodology also leads to higher ε and smaller σ values than approach A. This trend of parameter values obtained through approach B reflects the higher weight assigned to more stable conformers in the parameter fitting procedure. A higher approximation between the molecules is possible in more stable conformers due to smaller steric hindrances.

Table 3 – Top-down-based Mie potential parameters within the SAFT framework available in the literature and bottom-up-based Mie potential parameters obtained in this study from ab initio data by different methodologies and contexts for some substituted-methane compounds.

			Mie potential parameters[a,b,f]					
Molecule	Methodology[c]	Context[d]	m	σ/Å	λ_{rep}	λ_{att}	(ε/k_B)/K	Reference
tetrafluormethane	TD	SAFT framework	1.0000	4.3827	38.5075	6.0000	269.6989	[e]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.3246	17.6995	6.0000	218.8569	this work
		MP2/aug-cc-pVTZ (B)	1.0000	3.6898	12.0000	6.0000	275.6330	this work
fluoromethane	TD	SAFT framework	1.0000	3.9422	41.7173	6.0000	383.2007	[e]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.1530	13.8822	6.0000	259.8926	this work
		MP2/aug-cc-pVTZ (B)	1.0000	2.9662	12.0000	6.0000	965.1931	this work
chlorotrifluoromethane	TD	SAFT framework	1.0000	4.7189	37.2311	6.0000	355.1899	[e]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.7029	17.8595	6.0000	315.0594	this work
		MP2/aug-cc-pVTZ (B)	1.0000	3.8534	12.0000	6.0000	476.7379	this work
dichlorodifluoromethane	TD	SAFT framework	1.0000	5.0059	38.5420	6.0000	456.4715	[e]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.9122	18.2487	6.0000	462.8079	this work
		MP2/aug-cc-pVTZ (B)	1.0000	3.8447	12.0000	6.0000	840.5271	this work
dichlorofluoromethane	TD	SAFT framework	1.0000	4.8676	42.9811	6.0000	548.7515	[e]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.8666	17.5977	6.0000	493.5344	this work
		MP2/aug-cc-pVTZ (B)	1.0000	3.7152	12.0000	6.0000	1230.0140	this work
chlorodifluoromethane	TD	SAFT framework	1.0000	4.5956	45.8045	6.0000	454.8120	[e]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.6220	16.7608	6.0000	344.2269	this work
		MP2/aug-cc-pVTZ (B)	1.0000	3.6197	12.0000	6.0000	996.1710	this work

[a]Molecules are represented as spherical 1-site coarse-grained model (in the SAFT-VR Mie equation of state is equivalent to assigning the number of segments m equal to 1). [b]k_B is the Boltzmann constant. [c]TD and BU refer to the top-down and bottom-up methodologies, respectively, while (S) refers to the supermolecular methods. [d](A) and (B) indicate that approaches A and B objective functions, respectively, were minimized in the parameter fitting procedure. [e]Professor Erich A. Müller (personal communication, June 3rd, 2021), based on published works [106, 107]. [f]The best results for fitting the parameters of approach (B) were observed by setting the λ_{rep} value equal to 12.

The values of the Mie potential parameter λ^{rep} obtained by the bottom-up methodology are much smaller than those obtained by the top-down methodology. Dufal *et al.* [103] showed that different sets of Mie potential parameters can lead to similar results when predicting thermophysical properties, indicating that the Mie potential parameters in SAFT-VR Mie equa-

tion of state are not entirely independent. Comparing the Mie potential parameters obtained by bottom-up and top-down methodologies, the parameters obtained by the bottom-up methodology are more reliable since they are obtained to replicate interaction energy data between molecules. Therefore, within a set of Mie potential parameters that accurately predict thermophysical properties for any molecule, the set with parameter values closest to those obtained through the bottom-up approach should be the most reliable.

Fig. 17 graphically represents the interaction energies obtained by *ab initio* calculations and the Mie potential energy curves using top-down and bottom-based Mie potential parameters for some substituted-methane compounds of this study. From Fig. 17, the interaction energy curves obtained using top-down-based Mie potential parameters also seem to approximate the average of the *ab initio* data considered in obtaining the parameters of the bottom-up approach. The coverage area of *ab initio* data represented by the gray shaded areas in Fig. 17 highlights the observation previously made.

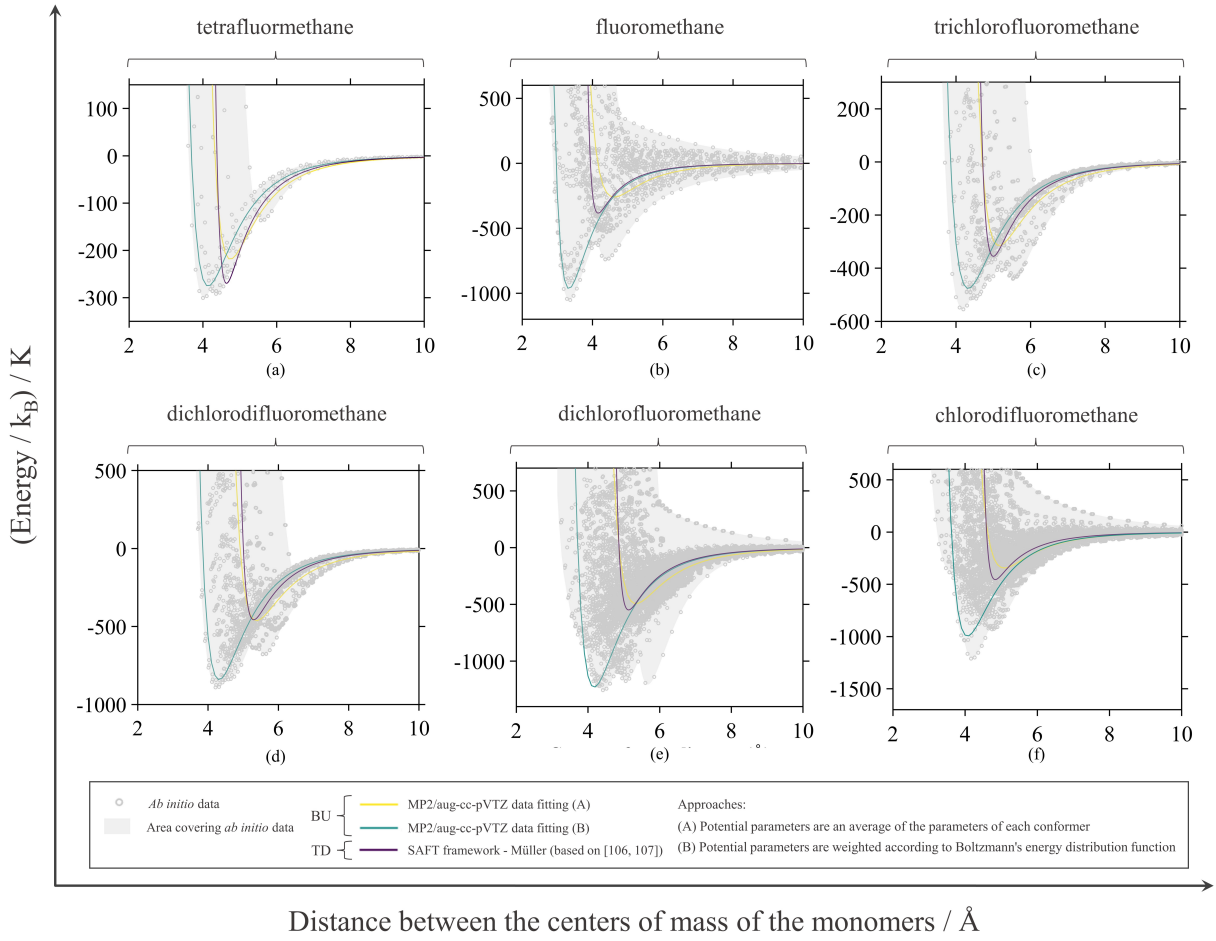

Figure 17 – Comparison of Mie potential energy curves obtained using parameters from top-down (TD) and bottom-up (BU) methodologies and *ab initio* interaction energy data obtained via supermolecular for some substituted-methane compounds of this study. Results for (a) tetrafluoromethane, (b) fluoromethane, (c) chlorotrifluoromethane, (d) dichlorodifluoromethane, (e) dichlorofluoromethane, and (f) chlorodifluoromethane.

The top-down and bottom-up based Mie potential parameters obtained in this study were also used in the SAFT-VR Mie equation of state and through molecular dynamics simulations to predict thermophysical properties. The prediction results were compared with NIST reference data by computing AARD% values. Fig. 18 presents the AARD% values for density (ρ), isochoric heat capacity (c_v), isobaric heat capacity (c_p), and speed of sound (c_{sound}) computed for each substituted-methane compound studied. The information presented in Fig. 18 is also available alternatively in Table D7 in Appendix D. The dataset used to compute thermophysical properties comprises of a series of temperature and pressure conditions.

The AARD% values presented in Fig. 18 demonstrate that the results obtained by the SAFT-VR Mie equation of state molecular dynamics simulations are consistent. The lowest values of AARD% were obtained again using top-down-based Mie potential parameters. When comparing approaches A and B of the bottom-up approach to determine Mie potential parameters, approach A presented significantly lower AARD% values than approach B, mainly for ρ and c_{sound}. Although approach A had the lowest AARD% values for ρ and c_{sound}, they were still quite high (in the range of tens of percent) for non-tetrahedral molecules such as dichlorodifluoromethane, fluoromethane, chlorotrifluoromethane, chlorodifluoromethane, and dichlorofluoromethane.

The tetrahedral geometry of tetrafluoromethane makes the spherical 1-site coarse-grained model a suitable representation for this molecule. This is supported by the lower AARD% values observed for tetrafluoromethane compared to other substituted-methane compounds that were studied. In the case of tetrafluoromethane, both bottom-up-based Mie potential parameters from approach A and top-down-based Mie potential parameters from the SAFT framework resulted in similar AARD% values when calculating c_v and c_p. For non-tetrahedral molecules from the set of substituted-methane compounds, both top-down and bottom-up-based Mie potential parameters of approach A were unable to predict c_v and c_p well, with values of AARD% exceeding more than thousand percent.

The previously reported observations on the obtained AARD% values imply that the spherical 1-site coarse-grained model used to represent non-tetrahedral substituted-methane compounds may be inappropriate. Non-tetrahedral substituted-methane compounds are strongly dependent on molecular orientation, a fact previously evidenced by the larger gray shaded areas representing the *ab initio* data in Fig. 17. As there are purely repulsive interaction energy curves for non-tetrahedral substituted-methane compounds, the Mie potential may also be unsuitable to represent these types of compounds. These considerations may explain the higher AARD% values obtained for non-tetrahedral substituted-methane compounds.

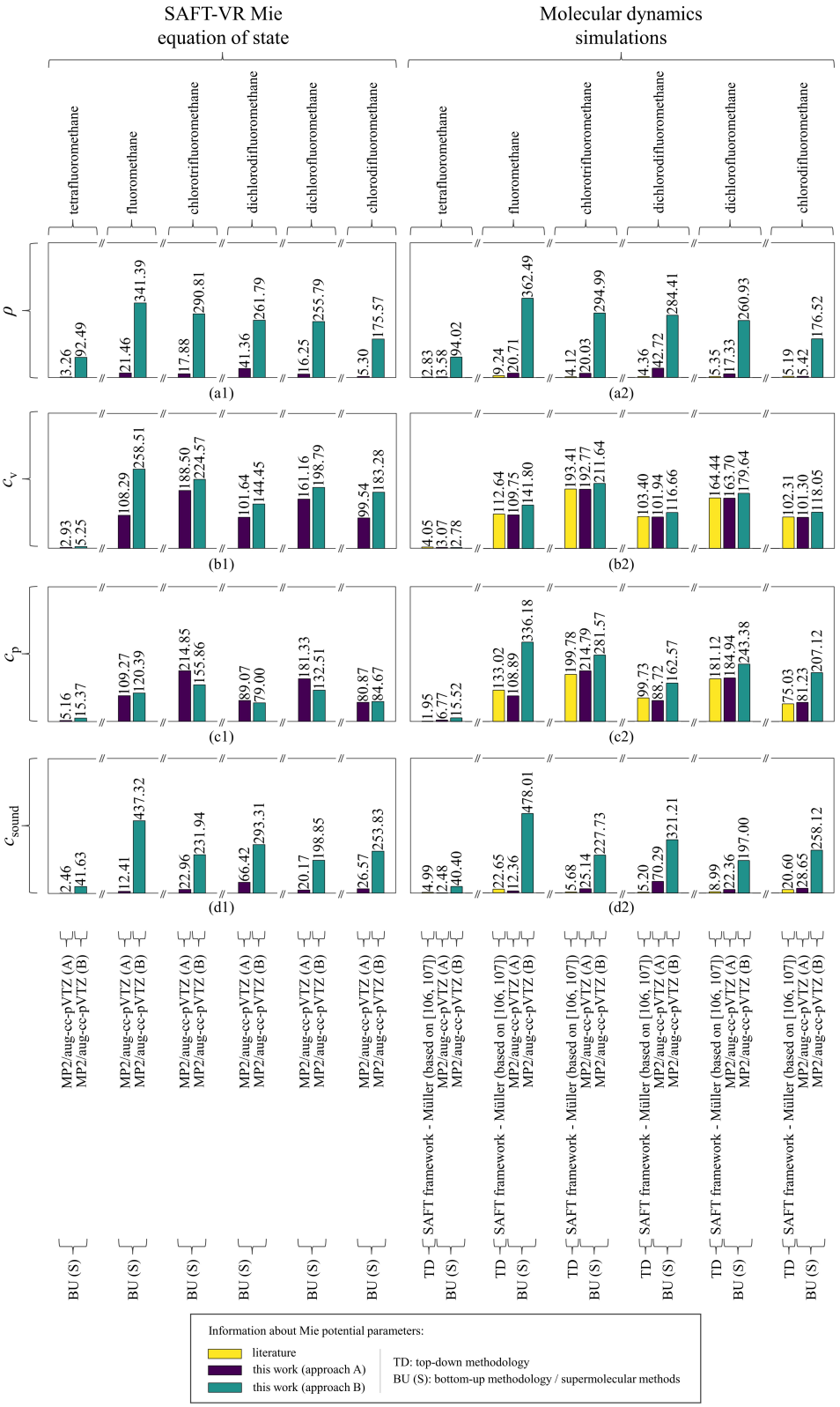

Figure 18 – AARD% values for thermophysical properties for some substituted-methane compounds (tetrafluoromethane, fluoromethane, chlorotrifluoromethane, dichlorodifluoromethane, dichlorofluoromethane, and chlorodifluoromethane) using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state and molecular dynamics simulations. Results for (a1) ρ, (b1) c_v, (c1) c_p, and (d1) c_{sound} by SAFT-VR Mie equation of state. Results for (a2) ρ, (b2) c_v, (c2) c_p, and (d2) c_{sound} by molecular dynamics simulations.

The critical point of substituted-methane compounds was also computed using the SAFT-VR Mie equation of state. The values for percentage deviations from NIST reference data for top-down and approach A bottom-up based parameters are presented in Fig. 19. The information presented in Fig. 19 is also available alternatively in Table D8 in Appendix D. The results presented in Fig. 19 indicate that both top-down and approach A bottom-up-based Mie potential parameters generated critical point close to NIST reference data. The most significant percentage deviations from NIST reference data are observed for non-tetrahedral substituted-methane compounds due to the inadequacy of the proposed model to molecular geometry. These findings align with previous observations regarding the AARD% values.

Due to the greater adequacy of the proposed model, Fig. 20 presents graphs that represent the capacity to predict thermophysical properties from NIST reference data for tetrafluoromethane using approach A Mie potential parameters in SAFT-VR Mie equation of state and molecular dynamics simulations. As for methane, the results in Fig. 20 demonstrate that the spherical 1-site coarse-grained model is suitable for representing tetrafluoromethane since a good representation of NIST reference data was obtained. For non-tetrahedral substituted-methane compounds, a similar presentation done for tetrafluoromethane can be found in the Appendix E of this study.

The SAFT-VR Mie equation of state is used to compute saturation properties of the substituted-methane compounds considered. In Fig. 21 the values of AARD% for saturation pressure (p_{sat}) and enthalpy of vaporization (ΔH_{vap}) considering some bottom-up and top-down-based Mie potential parameters. The temperature and pressure ranges and the number of NIST reference data considered in the calculations of the AARD% values presented in Fig. 21 can be found in Table C2 of the Appendix C. The information presented in Fig. 21 is also available alternatively in Table D9 in Appendix D. As it better fits the spherical 1-site coarse-grained model, the lowest AARD% values for saturation properties were observed for tetrafluoromethane. For the other non-tetrahedral molecules, higher values of AARD% are observed for saturation properties, something analogous to the findings for thermophysical properties and critical point. Previously in this work, it was observed that for methane, the reference values for p_{sat} considered significantly impact the calculation of AARD%. Thus, graphs were constructed to represent the predictive capacity of saturation properties for Mie potential parameters obtained by the bottom-up methodology through the MP2/aug-cc-pVTZ theory level and approach A, as shown in Fig. 22.

From Fig. 22 a good predictive capacity for saturation properties of tetrafluoromethane, and a reasonable capacity for non-tetrahedral molecules are observed. As the model lacks accuracy in representing the critical point, the reduction in temperature leads to an increase in the deviation of the computed thermophysical properties in relation to NIST reference data when using Mie potential parameters obtained by the bottom-up methodology through the MP2/aug-cc-pVTZ theory level and approach A.

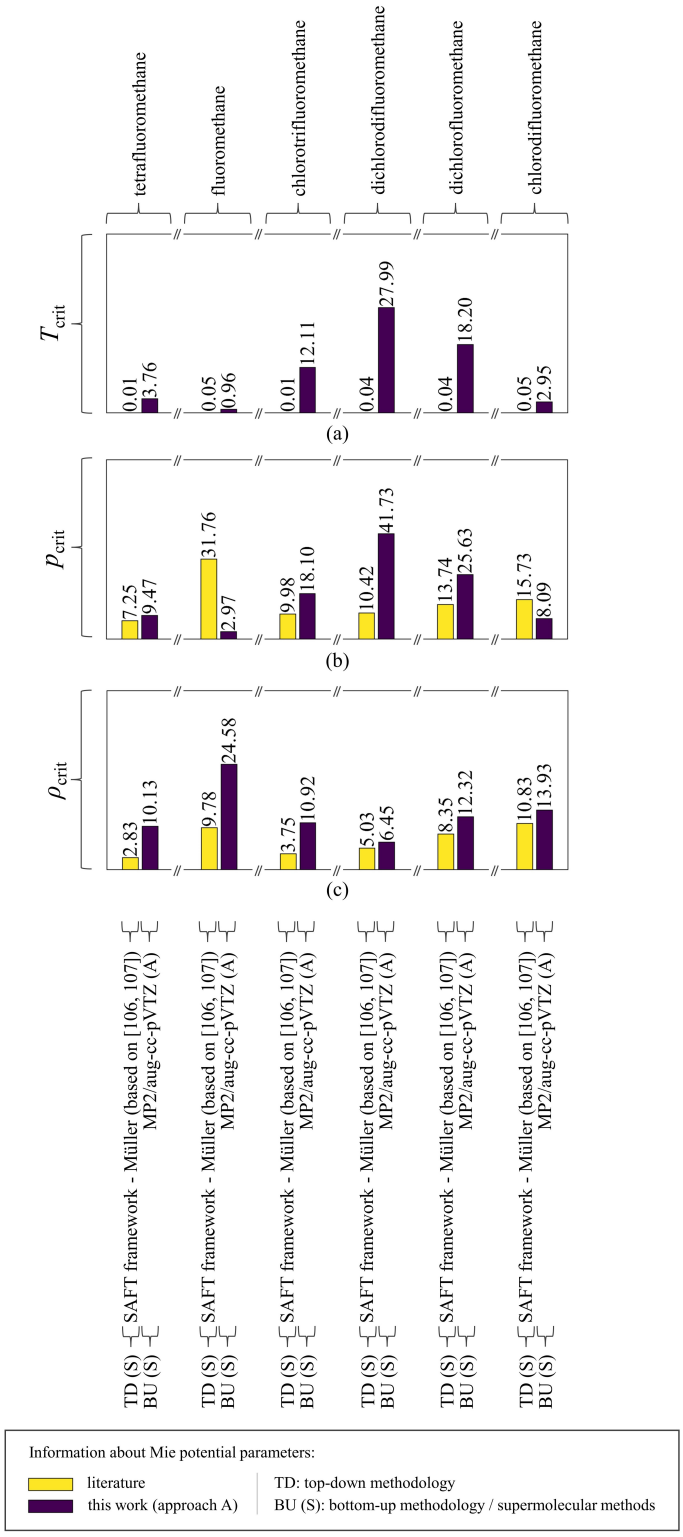

Figure 19 – Percentage deviation of the calculated critical point from NIST reference data using top-down and approach A bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state for some substituted-methane compounds (tetrafluoromethane, fluoromethane, chlorotrifluoromethane, dichlorodifluoromethane, dichlorofluoromethane, and chlorodifluoromethane). Results for (a) T_{crit}, (b) p_{crit}, and (c) ρ_{crit}.

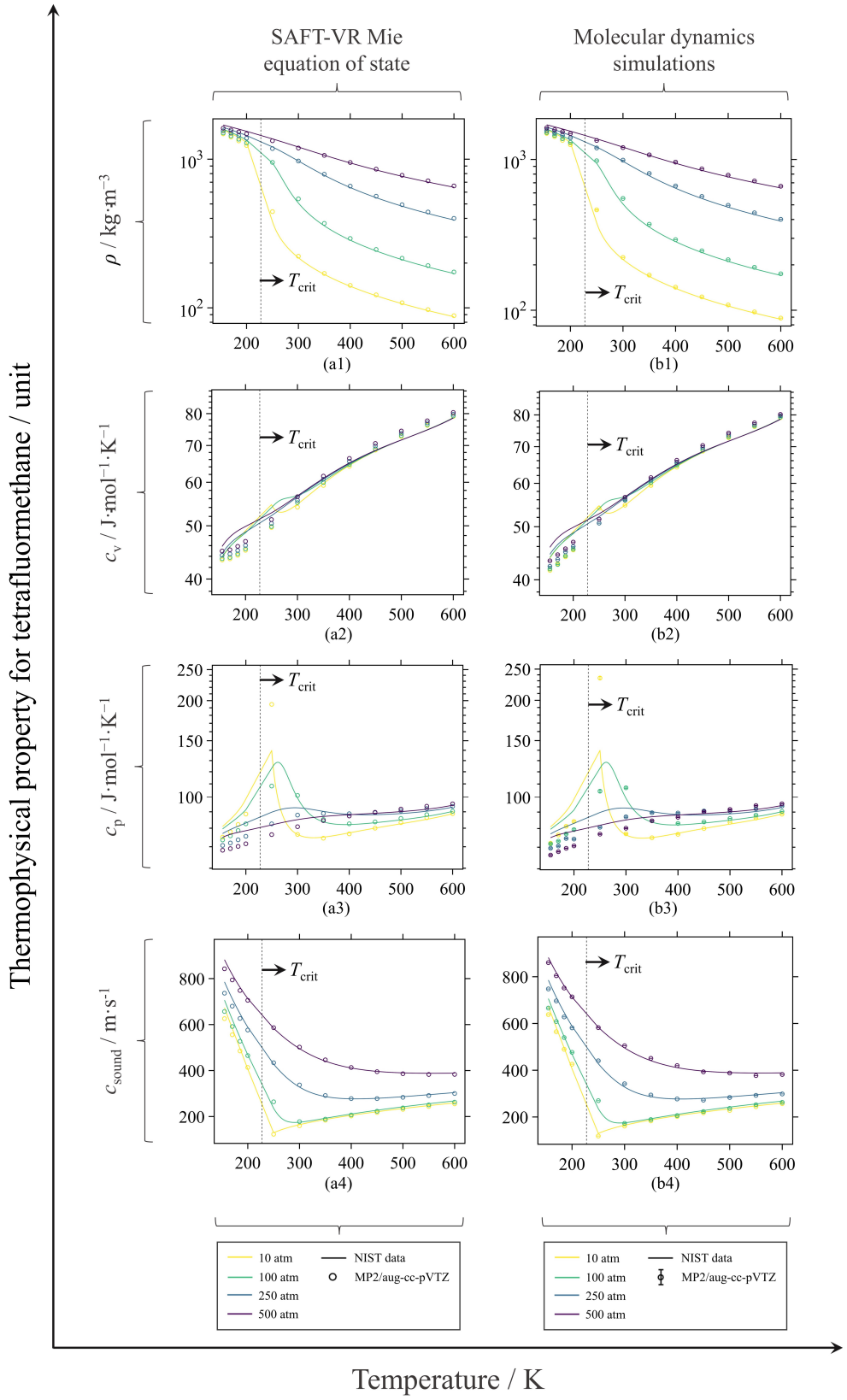

Figure 20 – Predicted thermophysical properties for tetrafluoromethane using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Prediction of (a1) ρ, (b1) c_v, (c1) c_p, and (d1) c_{sound} by SAFT-VR Mie equation of state. Prediction of (a2) ρ, (b2) c_v, (c2) c_p, and (d2) c_{sound} by molecular dynamics simulations.

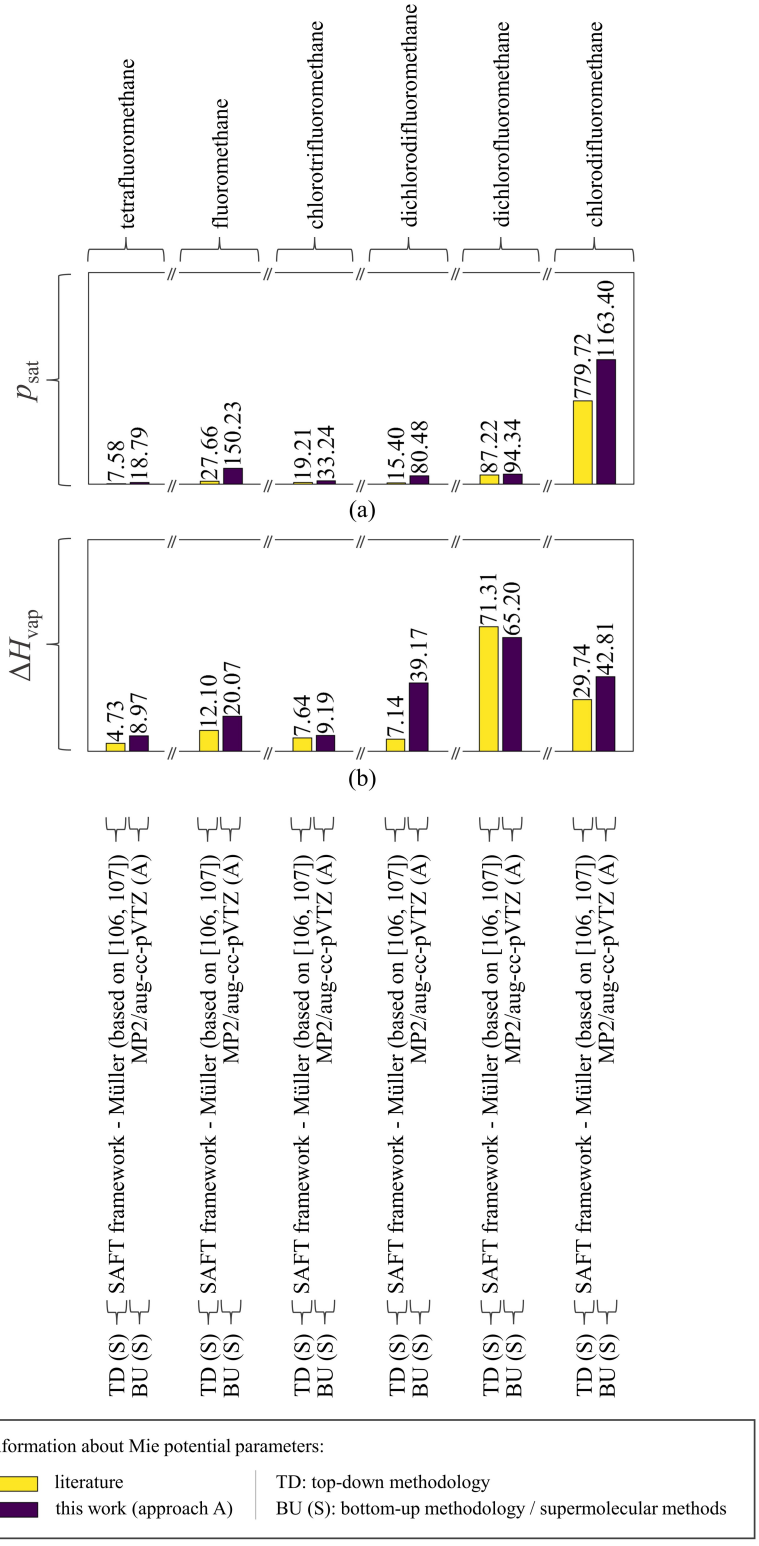

Figure 21 – AARD% values for saturation properties for some substituted-methane compounds using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state. Results of (a) p_{sat} and (b) ΔH_{vap} for tetrafluoromethane, fluoromethane, chlorotrifluoromethane, dichlorodifluoromethane, dichlorofluoromethane, and chlorodifluoromethane.

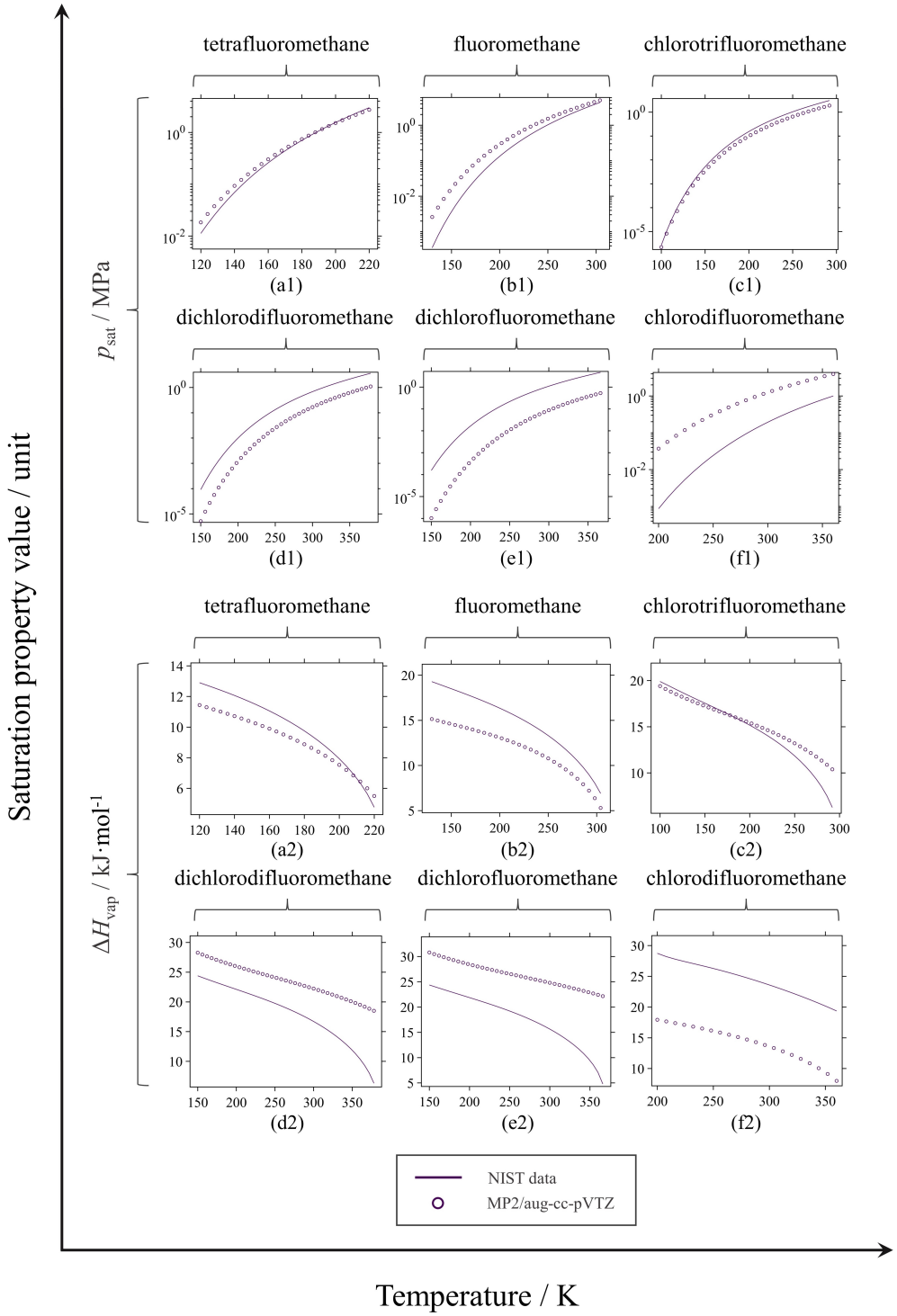

Figure 22 – Predicted saturation properties for some substituted-methane compounds using bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies by approach A in the SAFT-VR Mie equation of state. Results of p_{sat} for (a1) tetrafluoromethane, (b1) fluoromethane, (c1) chlorotrifluoromethane, (d1) dichlorodifluoromethane, (e1) dichlorofluoromethane, and (f1) chlorodifluoromethane. Results of ΔH_{vap} for (a2) tetrafluoromethane, (b2) fluoromethane, (c2) chlorotrifluoromethane, (d2) dichlorodifluoromethane, (e2) dichlorofluoromethane, and (f2) chlorodifluoromethane.

3.3 Third group: molecules of different chemical classes

The calculations performed in Sections 3.1 and 3.2 of this study focused on obtaining Mie potential parameters from *ab initio* calculations to make the SAFT-VR Mie equation of state purely predictive for thermophysical properties, critical point, and saturation properties of some noble gases, methane, and some substituted methane compounds. Obtaining Mie potential parameters through *ab initio* data of interaction energy curves for noble gases becomes easier since they are monoatomic and isotropic. In contrast, methane and substituted-methane compounds are anisotropic, which results in different interaction energy curves for dimers of different conformations. Obtaining the Mie potential parameters for methane and substituted-methane compounds was facilitated by the existence of previously studied dimer configurations investigated in the work of Chao *et. al.* [92]. Now, in this Section of this study, the discussion is extended on how to generate dimers of molecules with geometry different from those previously studied to obtain Mie potential parameters through *ab initio* calculations to configure a purely predictive character in the SAFT-VR Mie equation of state.

Similar to methane and substituted-methane compounds, a set of 15 isolated molecules had their chemical structure optimized at the MP2/aug-cc-pVQZ theory level. The studied molecules are hydrogen, nitrogen, oxygen, fluorine, carbon monoxide, carbon dioxide, ethane, propane, *n*-butane, isobutane, ethene, propene, propyne, cyclopropane, and benzene. Molecules with optimized geometry were used to build dimers in different conformations according to the assumptions reported for the third group of molecules presented in Fig. 3(c) presented in Section 2.2.2 of this study.

Due to the good results obtained for noble gases, methane, and substituted-methane compounds and the computational cost, the interaction energy curves of the dimers were obtained by supermolecular methods at the MP2/aug-cc-pVTZ theory level. The set of 15 studied molecules was also modeled using a spherical 1-site coarse-grained model based on the Mie potential. The Mie potential parameters σ, ε, and λ^{rep} were obtained by approach A of the bottom-up methodology for all studied molecules. Approach A was chosen because it generates bottom-up methodology parameters closer to those found in the literature for the top-down methodology within the SAFT framework in the groups of molecules previously studied. Approach A indirectly considers the spatial dependence in the Mie potential parameters obtained by the bottom-up methodology. Table 4 presents the top-down and bottom-up based Mie potential parameters obtained. Fig. 23 complements the results of Table 4 presenting the Mie potential energy curves obtained using parameters of approach A of the bottom-up methodology and the interaction energy curves obtained by *ab initio* calculations.

Table 4 – Top-down-based Mie potential parameters within the SAFT framework available in the literature and bottom-up-based Mie potential parameters obtained in this study from ab initio data by different methodologies for a group of 15 molecules of different chemical classes compounds.

			Mie potential parameters[a,b]					
Molecule	Methodology[c]	Context[d]	m	$\sigma/$	λ_{rep}	λ_{att}	$(\varepsilon/k_B)/K$	Reference
hydrogen	TD	SAFT framework	1.0000	3.1586	7.8130	6.0000	18.3550	[99]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	3.1995	8.2660	6.0000	21.6242	this work
nitrogen	TD	SAFT framework	1.4214	3.1760	9.8749	6.0000	72.4380	[103]
		SAFT framework	1.0000	3.6530	20.0200	6.0000	122.8500	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	3.5137	12.3709	6.0000	152.5762	this work
oxygen	TD	SAFT framework	1.4283	2.9671	8.9218	6.0000	81.4760	[103]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	3.4915	14.6902	6.0000	148.2514	this work
fluorine	TD	SAFT framework	1.5094	2.8118	9.9255	6.0000	80.8030	[103]
		SAFT framework	1.3211	2.9554	11.6060	6.0000	96.2680	[87]
		SAFT framework	1.0000	3.3186	21.9600	6.0000	145.1700	[107]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	3.3561	13.8871	6.0000	96.0516	this work
carbon monoxide	TD	SAFT framework	1.5556	3.0928	9.7420	6.0000	72.1100	[103]
		SAFT framework	1.0000	3.6870	21.4900	6.0000	132.8300	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	3.5771	13.4940	6.0000	120.4417	this work
carbon dioxide	TD	SAFT framework	1.6936	3.0465	18.0670	6.0000	235.7300	[103]
		SAFT framework	2.0000	2.8480	14.6500	6.0000	194.9400	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	3.6965	15.2710	6.0000	298.9670	this work
ethane	TD	SAFT framework	1.7230	3.4763	10.1210	6.0000	164.2700	[103]
		SAFT framework	1.4373	3.7257	12.4000	6.0000	206.1200	[87]
		SAFT framework	1.0000	4.3490	27.3000	6.0000	330.2500	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.2470	16.4039	6.0000	297.7097	this work
propane	TD	SAFT framework	1.8068	3.7943	12.1060	6.0000	221.9600	[103]
		SAFT framework	1.6845	3.9056	13.0060	6.0000	239.8900	[87]
		SAFT framework	1.0000	4.8710	34.2900	6.0000	426.0800	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.7931	19.5709	6.0000	390.0400	this work
n-butane	TD	SAFT framework	1.6791	4.2476	15.4530	6.0000	306.5200	[103]
		SAFT framework	1.8514	4.0887	13.6500	6.0000	273.6400	[87]
		SAFT framework	2.0000	3.9610	13.2900	6.0000	256.3600	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	5.2170	21.3595	6.0000	474.5755	this work
isobutane	TD	SAFT framework	1.7186	4.2177	14.6120	6.0000	281.1200	[103]
		SAFT framework	2.0000	3.9740	12.9400	6.0000	241.5700	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	5.1661	19.4316	6.0000	462.1039	this work
ethene	TD	SAFT framework	1.7972	3.2991	9.6463	6.0000	142.640	[103]
		SAFT framework	1.0000	4.1800	25.6200	6.0000	299.4900	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.1369	25.9756	6.0000	299.1178	this work
propene	TD	SAFT framework	2.0060	3.5392	10.6430	6.0000	190.1300	[103]
		SAFT framework	1.0000	4.7210	33.6500	6.0000	417.6000	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.3794	18.0501	6.0000	426.8011	this work
propyne	TD	SAFT framework	2.3038	3.2076	10.6670	6.0000	200.1100	[103]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.4088	16.9949	6.0000	461.6010	this work
cyclopropane	TD	SAFT framework	1.8769	3.4830	10.5070	6.0000	213.1100	[103]
		SAFT framework	1.0000	4.5110	31.1600	6.0000	447.9000	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.4508	17.3623	6.0000	431.4277	this work
benzene	TD	SAFT framework	2.2785	3.7806	11.5940	6.0000	297.5300	[103]
		SAFT framework	1.9163	4.0549	14.7980	6.0000	372.5900	[87]
		SAFT framework	2.0000	3.9780	14.2300	6.0000	353.9300	[104]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.0000	4.9840	19.8237	6.0000	975.8040	this work

[a] m refers to the number of segments considered to represent the molecules in the SAFT-VR Mie equation of state. [b] k_B is the Boltzmann constant. [c] TD and BU refer to the top-down and bottom-up methodologies, respectively, while (S) refers to the supermolecular methods. [d] (A) indicates that the objective function of approach A was minimized in the parameter tuning procedure.

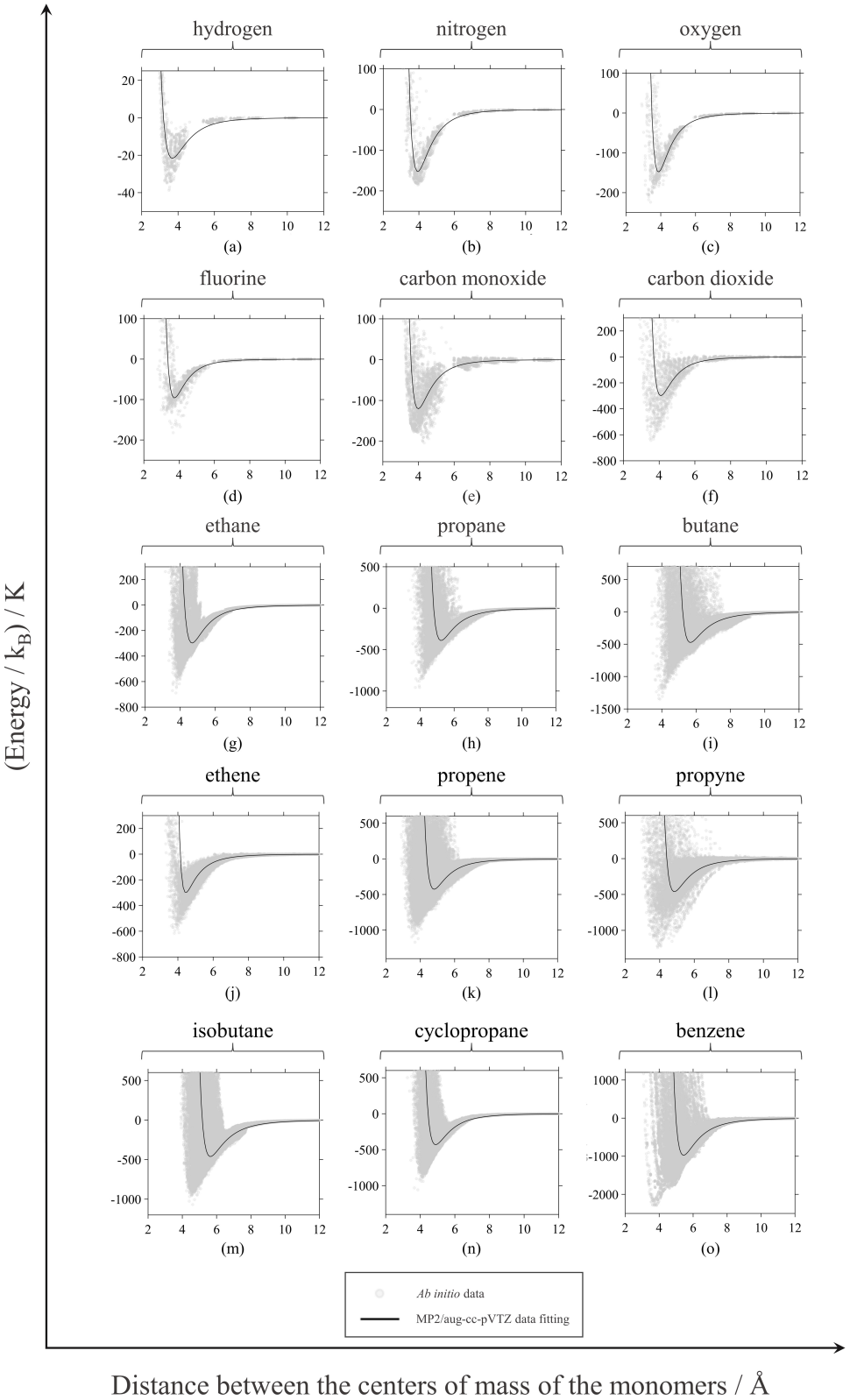

Figure 23 – Comparison of Mie potential energy curves obtained using parameters from bottom-up methodology and *ab initio* interaction energy data obtained via supermolecular methods for a group of fifteen molecules of different chemical classes compounds. Results for (a) hydrogen, (b) nitrogen, (c) oxygen, (d) fluorine, (e) carbon monoxide, (f) carbon dioxide, (g) ethane, (h) propane, (i) *n*-butane, (j) ethene, (k) propene, (l) propyne, (m) isobutane, (n) cyclopropane, (o) benzene.

Table 4 demonstrates that not all sets of top-down-based Mie potential parameters available in the literature within the SAFT framework were obtained considering a spherical 1-site coarse-grained (*i.e.*, considering a single spherical segment $m = 1$). This indicates that the spherical 1-site coarse-grained model may be unsuitable to represent thermophysical properties and critical point for some of the set of 15 studied molecules. Some references for top-down-based Mie potential parameters, however, considered a single spherical segment for describing the molecule. For these previous cases, the values of the Mie potential parameters σ and ε obtained by approach A of the bottom-up methodology were close to those found in the literature for the top-down methodology.

These results entail an important consistency check with the results previously observed for methane and substituted-methane compounds. Therefore, the top-down-based Mie potential parameters that considered a single spherical segment seem to approximate the average of the *ab initio* data considered in obtaining the parameters of the bottom-up approach. Once again, the values of the Mie potential parameter λ^{rep} proved to be very different when comparing the ones obtained by top-down and bottom-up methodologies. The bottom-up methodology underestimates the values of λ^{rep} obtained by the top-down methodology.

The top-down and bottom-up-based Mie potential parameters were used in the SAFT-VR Mie equation of state to predict the thermophysical properties for the set of 15 studied molecules. The thermophysical properties evaluated were density (ρ), isochoric heat capacity (c_{v}), isobaric heat capacity (c_{p}), speed of sound (c_{sound}), and Joule-Thomson coefficient (μ_{JT}). The predicted data were compared with NIST reference values. Fig. 24 compares the AARD% values using the top-down and bottom-up based Mie potential parameters for the set of 15 studied molecules. The AARD% values presented in Fig. 24 were calculated considering NIST reference data listed in Table C1 found in Appendix C. The information presented in Fig. 24 is also available alternatively in Table D10 in Appendix D.

Fig. 24 indicates that the lowest AARD% values were when using top-down-based Mie potential parameters, which are directly adjusted to reproduce experimental data. For certain references of top-down-based Mie potential, the number of the spherical segments (*i.e.*, m) was also optimized in the parameter fitting procedure. For these cases, there was an improvement in the AARD% values obtained compared to fitted considering a single spherical segment.

When bottom-up Mie potential parameters were used, the AARD% values obtained for density (ρ), heat capacity at constant volume (c_{v}), heat capacity at constant pressure (c_{p}), and speed of sound (c_{sound}) were low for almost all molecules, with the exception of fluorine, propene, propyne, cyclopropane and benzene. For these last five molecules, a spherical 1-site coarse-grained model may be inappropriate. For all molecules, the AARD% values obtained for the Joule-Thomson coefficient were high for both top-down and bottom-up-based Mie potential parameters, excluding some exceptions where the values were smaller than the order of tens of percent.

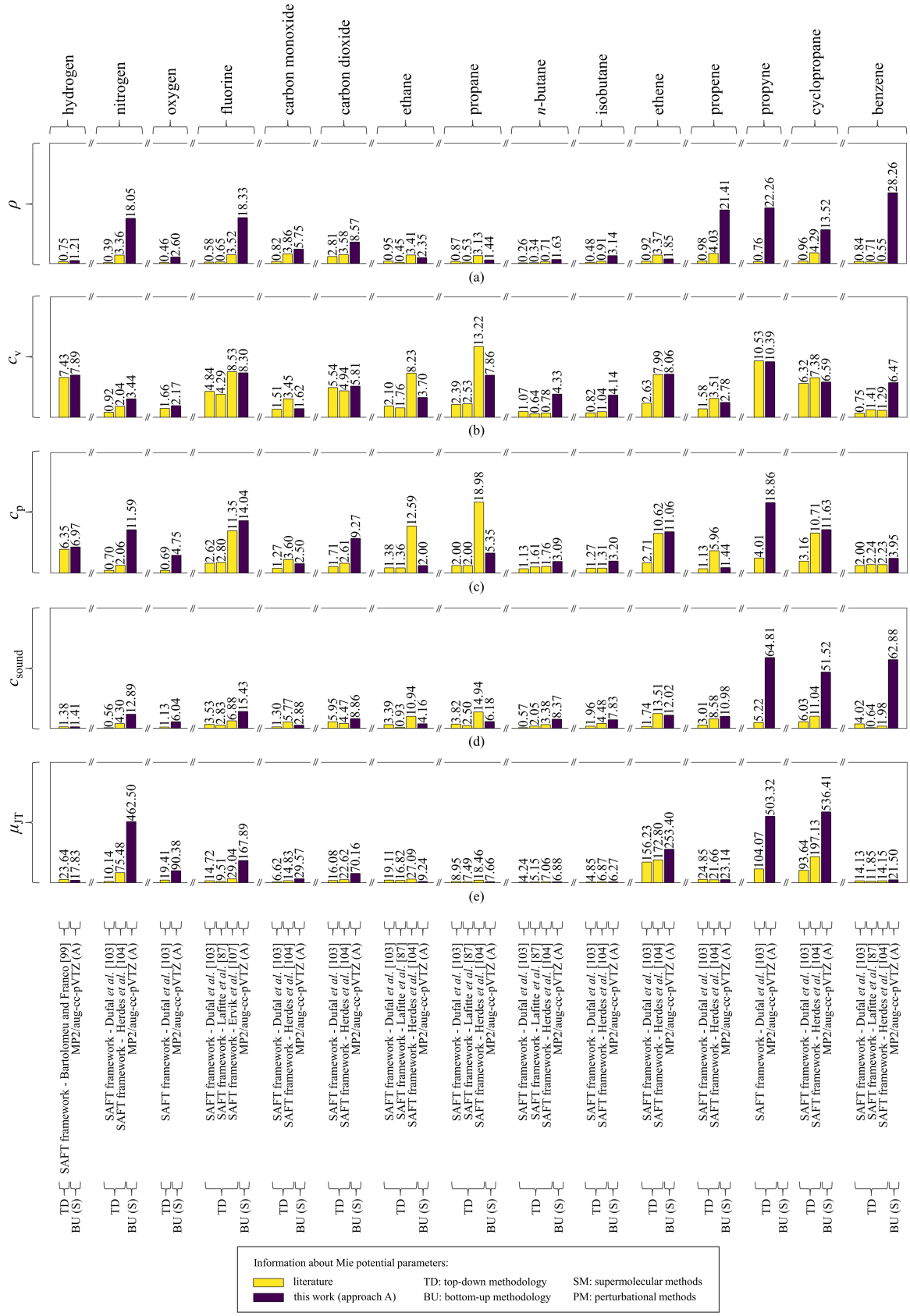

Figure 24 – AARD% values for thermophysical properties for 15 molecules of different chemical classes using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state. Results of (a) ρ, (b) c_v, (c) c_p, (d) c_{sound}, and (e) μ_{JT} for hydrogen, nitrogen, oxygen, fluorine, carbon monoxide, carbon dioxide, ethane, propane, *n*-butane, isobutane, ethene, propene, propyne, cyclopropane, and benzene.

The capacity to predict thermophysical properties using bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state was illustrated by plotting graphs that compare the predicted results with NIST reference data. In Fig. 25, the prediction results for hydrogen, nitrogen, oxygen, and fluorine are illustrated, while in Fig. 26, the results for carbon monoxide and carbon dioxide are depicted. For both Figs. 25 and 26, the capacity to predict thermophysical properties improves with increasing temperature, following the trend of the NIST reference data. In Fig. 25(d), the higher values of AARD% observed in Fig. 24 imply a lower capacity to predict the thermophysical properties of fluorine.

The results of predicting the thermophysical properties for some linear hydrocarbons (*i.e.*, methane, ethane, propane, and *n*-butane) are shown in Fig. 27. In Fig. 27 the predictive capacity for methane is that previously presented in this study by Fig. 14 in section 3.2.1, considering the Mie potential parameters obtained by approach A of the bottom-up methodology with MP2/aug-cc-pVTZ theory level energies. Fig. 27 demonstrates that the Mie potential parameters obtained by interaction energy curves with data from *ab initio* calculations can represent the trend of the NIST reference data very well. This fact is supported by the low AARD% values presented for these molecules in Fig. 24.

The prediction of thermophysical properties was also evaluated for branched hydrocarbons. Fig. 28 presents the predictive results of thermophysical properties for isobutane. Comparing the AARD% values observed for butane and isobutane in Fig. 24, Fig. 28 demonstrates that even with the spatial redistribution of the atoms in the butane molecule to generate isobutane, the predictive capacity remained good with the use of approach A bottom-up-based Mie potential parameters.

Another class of molecules whose capacity for predicting thermophysical properties has been evaluated is that of unsaturated hydrocarbons. Fig. 29 represents the predictive capabilities obtained for ethene, propene, and propyne. From Fig. 29, the best predictive results were observed for ethene, followed by propene and propyne. These observations are supported by the lower AARD% values obtained for ethene, followed by propene and propyne, respectively. For these molecules, the spherical 1-site coarse-grained model may be unsuitable since the number of the spherical segments is greater than one (*i.e.*, $m > 1$) in top-down-based Mie potential parameters found in the literature.

Finally, the capacity to predict thermophysical properties for cyclic hydrocarbons (both non-aromatic and aromatic) was also assessed. Cyclopropane (non-aromatic) and benzene (aromatic) were evaluated, with their results presented in Figs. 30 and 31, respectively. These figures show the largest discrepancies between the predictive model and NIST reference data, also reflected in the highest AARD% values for these compounds according to Fig. 24 data. For these molecules, the number of spherical segments considered is also greater than one (*i.e.*, $m > 1$) in top-down-based Mie potential parameters found in the literature. This suggests that the spherical 1-site coarse-grained model may also be unsuitable for these molecules.

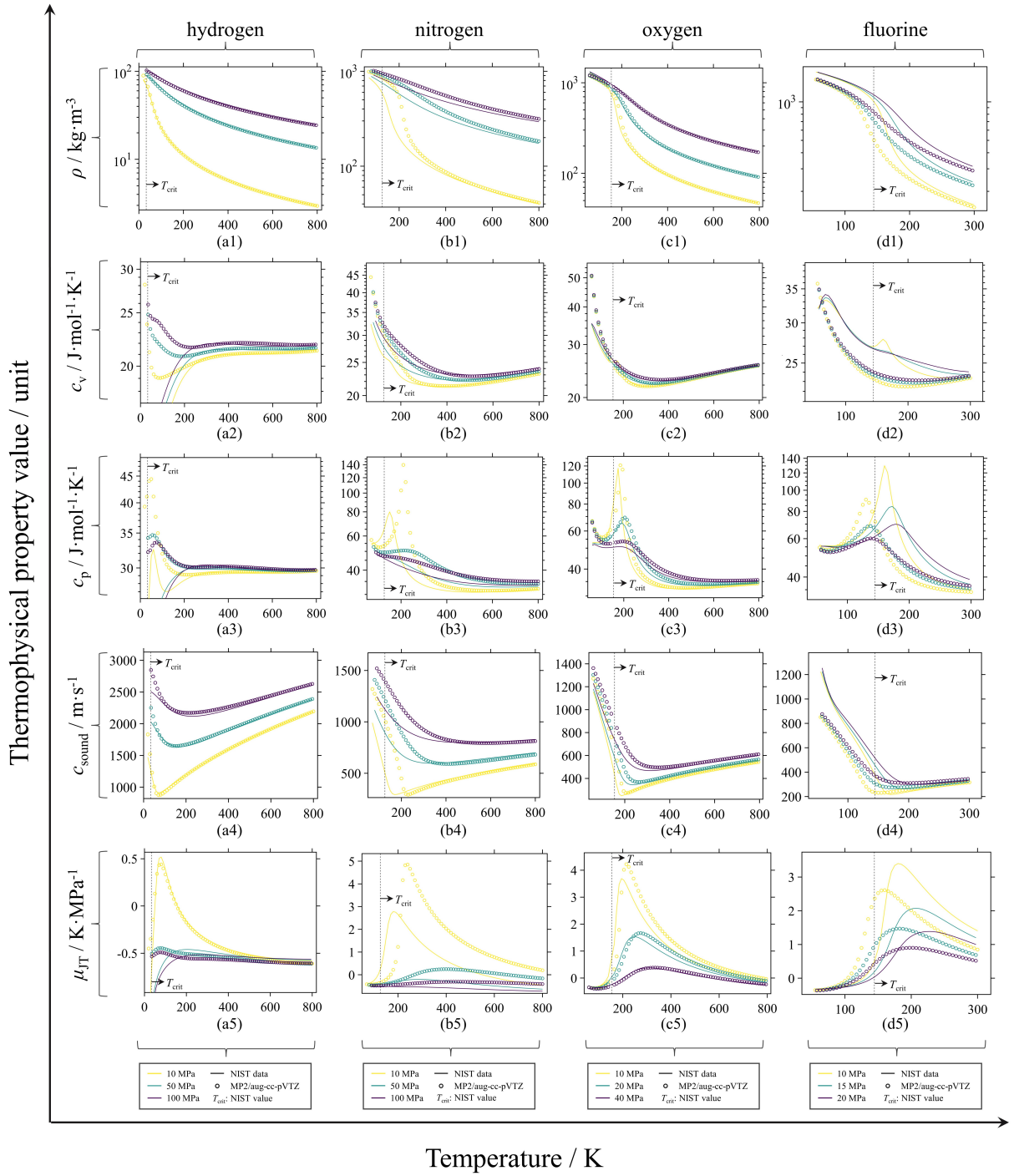

Figure 25 – Predicted thermophysical properties for hydrogen, nitrogen, oxygen, and fluorine using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Hydrogen results for (a1) ρ, (a2) c_v, (a3) c_p, (a4) c_{sound}, and (a5) μ_{JT}. Nitrogen results for (b1) ρ, (b2) c_v, (b3) c_p, (b4) c_{sound}, and (b5) μ_{JT}. Oxygen results for (c1) ρ, (c2) c_v, (c3) c_p, (c4) c_{sound}, and (c5) μ_{JT}. Fluorine results for (d1) ρ, (d2) c_v, (d3) c_p, (d4) c_{sound}, and (d5) μ_{JT}.

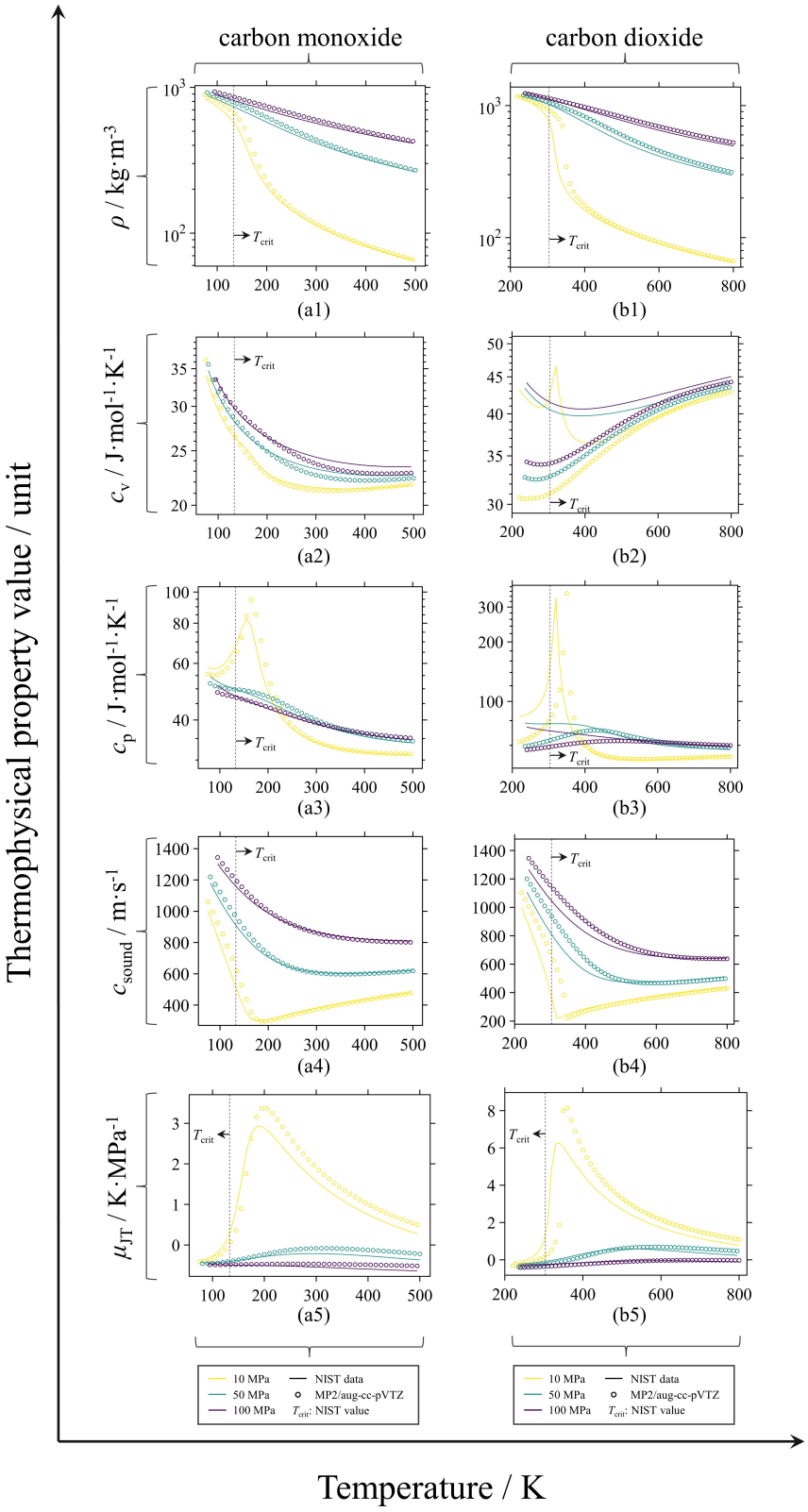

Figure 26 – Predicted thermophysical properties for carbon monoxide and carbon dioxide using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Carbon monoxide results for (a1) ρ, (a2) c_v, (a3) c_p, (a4) c_{sound}, and (a5) μ_{JT}. Carbon dioxide results for (b1) ρ, (b2) c_v, (b3) c_p, (b4) c_{sound}, and (b5) μ_{JT}.

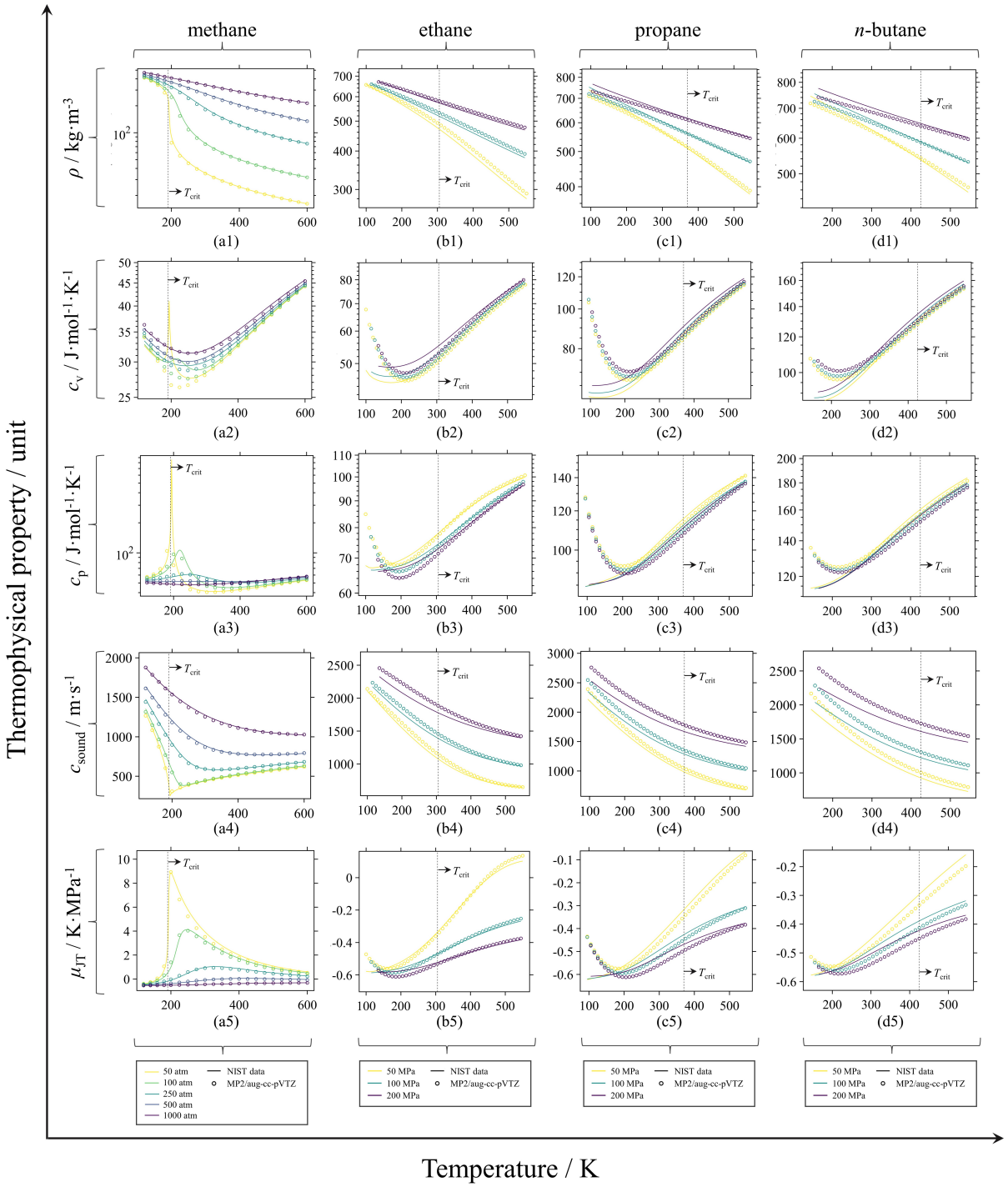

Figure 27 – Predicted thermophysical properties for some linear hydrocarbons using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Methane results for (a1) ρ, (a2) c_v, (a3) c_p, (a4) c_{sound}, and (a5) μ_{JT}. Ethane results for (b1) ρ, (b2) c_v, (b3) c_p, (b4) c_{sound}, and (b5) μ_{JT}. Propane results for (c1) ρ, (c2) c_v, (c3) c_p, (c4) c_{sound}, and (c5) μ_{JT}. *n*-Butane results for (d1) ρ, (d2) c_v, (d3) c_p, (d4) c_{sound}, and (d5) μ_{JT}.

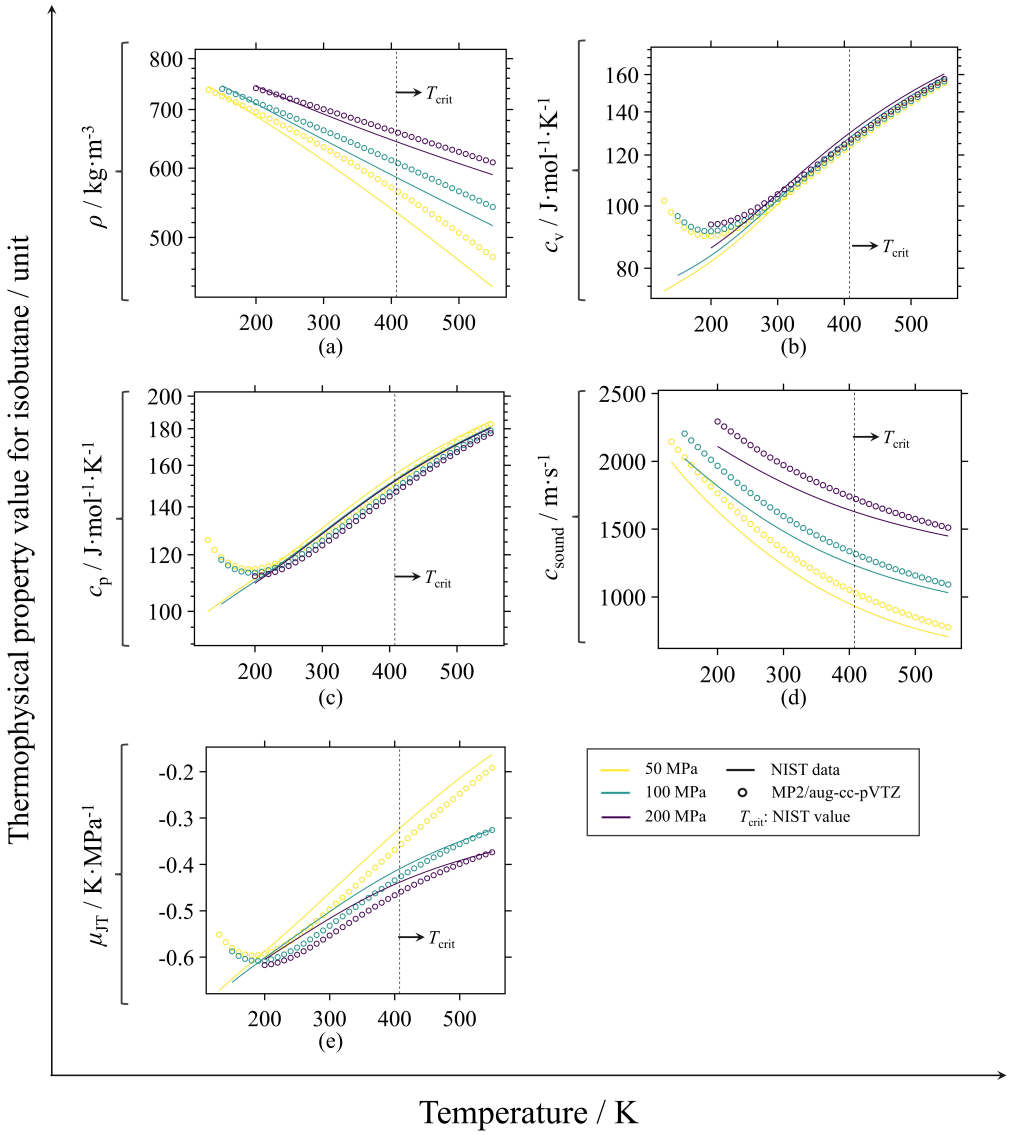

Figure 28 – Predicted thermophysical properties for a branched hydrocarbon using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Isobutane results for (a) ρ, (b) c_v, (c) c_p, (d) c_{sound}, and (e) μ_{JT}.

The top-down and bottom-up-based Mie potential parameters were also used to compute critical temperature (T_{crit}), pressure (p_{crit}), and density (ρ_{crit}) through the SAFT-VR Mie equation of state. In Fig. 32, the critical point values computed for each molecule in each different methodology are compared with NIST reference data through percentage deviation. Alternatively, the information presented in Fig. 32 is also available in Table D11 in Appendix D. As they are directly fitted to reproduce experimental data, the smallest percentage deviations observed to NIST reference data are for the top-down-based Mie potential parameters. The critical point predicted using bottom-up-based Mie potential parameters presented consistent results for molecules that the proposed spherical 1-site coarse-grained model was suitable, with deviations of tens of percent relative to the NIST reference data.

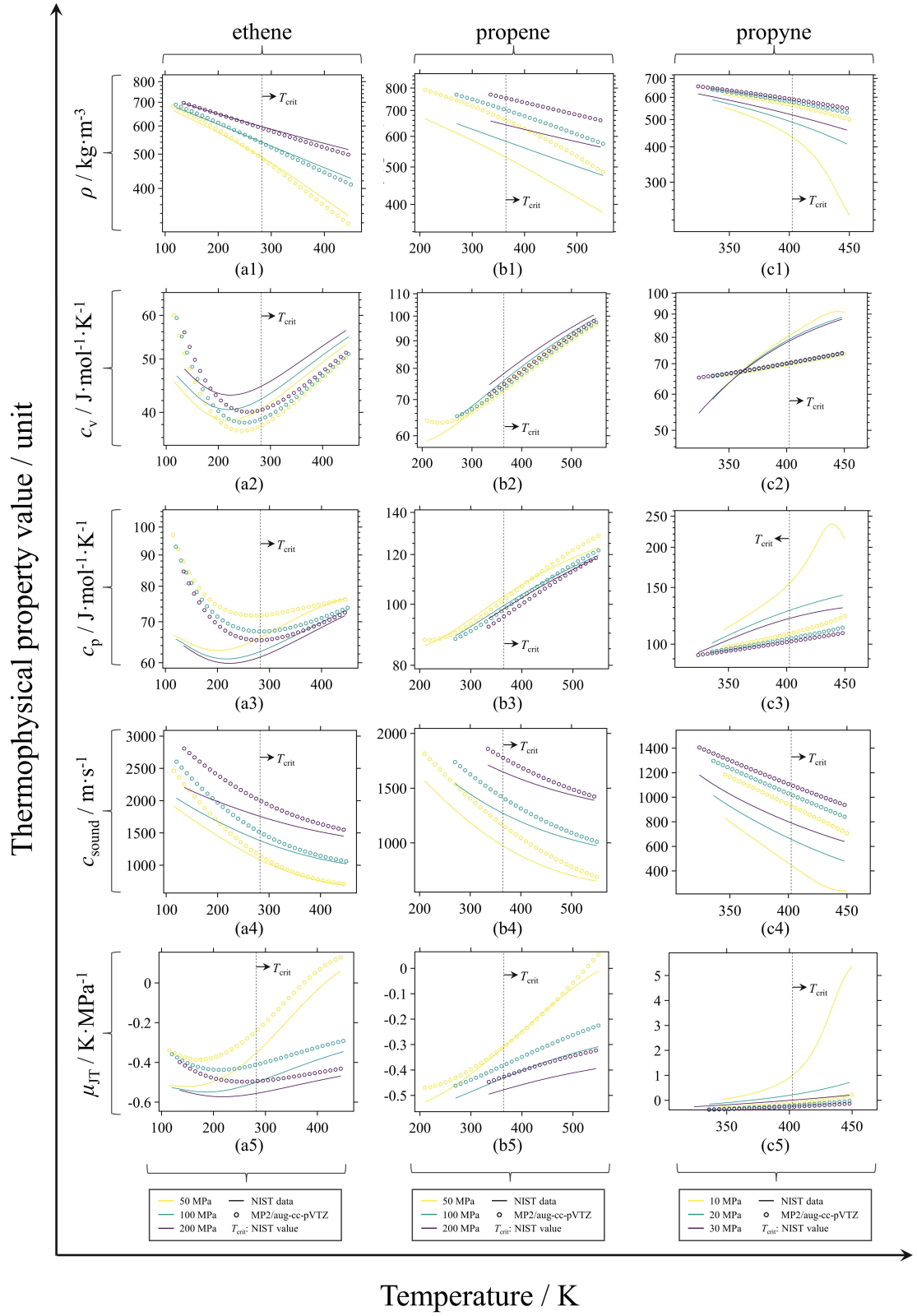

Figure 29 – Predicted thermophysical properties for some unsaturated hydrocarbons using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Ethene results for (a1) ρ, (a2) c_v, (a3) c_p, (a4) c_{sound}, and (a5) μ_{JT}. Propene results for (b1) ρ, (b2) c_v, (b3) c_p, (b4) c_{sound}, and (b5) μ_{JT}. Propyne results for (c1) ρ, (c2) c_v, (c3) c_p, (c4) c_{sound}, and (c5) μ_{JT}.

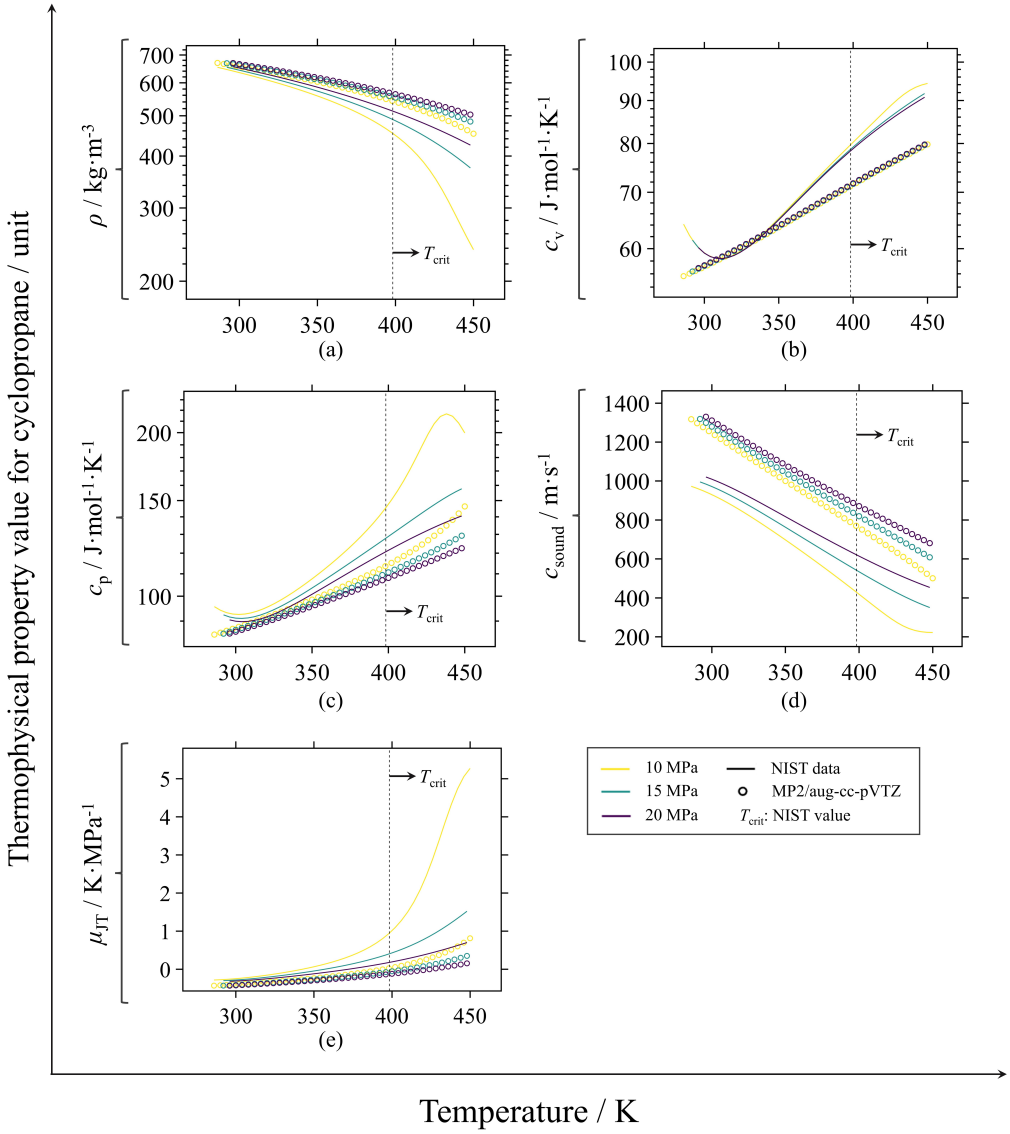

Figure 30 – Predicted thermophysical properties for a cyclic non-aromatic hydrocarbon using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Cyclopropane results for (a) ρ, (b) c_v, (c) c_p, (d) c_{sound}, and (e) μ_{JT}.

As a final analysis, saturation pressure (p_{sat}) and enthalpy of vaporization (ΔH_{vap} were also computed using some bottom-up and top-down-based Mie potential parameters through the SAFT-VR Mie equation of state. Fig. 33 demonstrates the AARD% values obtained for the fifteen studied molecules from different chemical classes. The AARD% values presented in Fig. 33 were calculated considering NIST reference data listed in Table C2 found in Appendix C. Alternatively, the information presented in Fig. 33 is also available in Table D12 in Appendix D. Because they are directly fitted to experimental data, the AARD% values presented for saturation properties using top-down-based Mie potential parameters were lower than when using bottom-down-based ones, as expected. Most of the AARD% values were in the order of tens of percent, with the exception of the values of p_{sat} for fluorine and ΔH_{vap} for nitrogen and benzene reaching higher values.

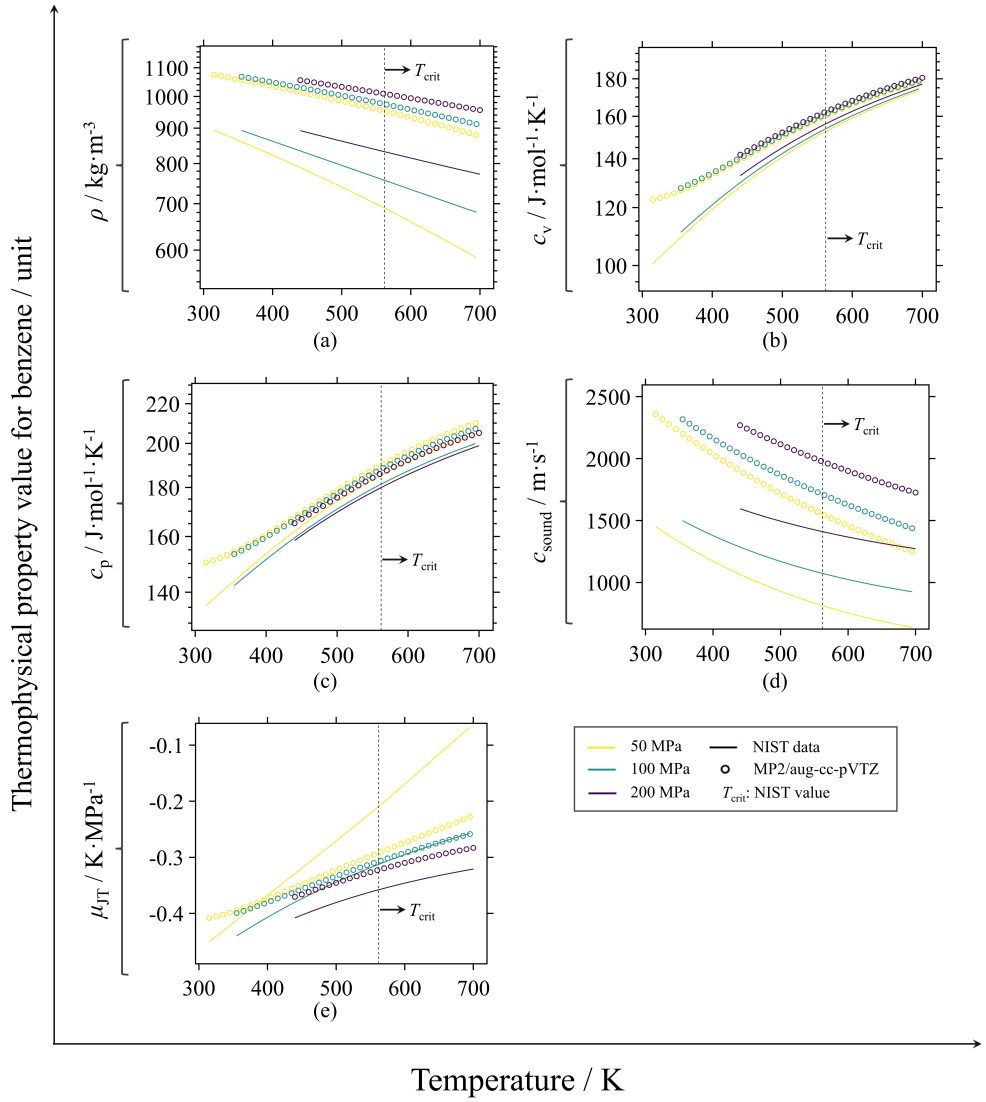

Figure 31 – Predicted thermophysical properties for a cyclic aromatic hydrocarbon using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Benzene results for (a) ρ, (b) c_v, (c) c_p, (d) c_{sound}, and (e) μ_{JT}.

In Fig. 34 the predictive capacity to reproduce saturation properties for Mie potential parameters obtained by the bottom-up methodology through the MP2/aug-cc-pVTZ approach A. Fig. 34 indicates that for oxygen, carbon monoxide, some alkanes (*i.e.*, ethane, propane and butane) and alkane derivatives (*i.e.*, isobutane, ethene) there were the best prediction capabilities for p_{sat}. From the graphs in Fig. 34 it can be seen that among the temperature ranges considered for calculating the saturation properties, the values of p_{sat} close to the lower limit are very low, so calculation errors in SAFT-VR Mie significantly impact the AARD% calculation.

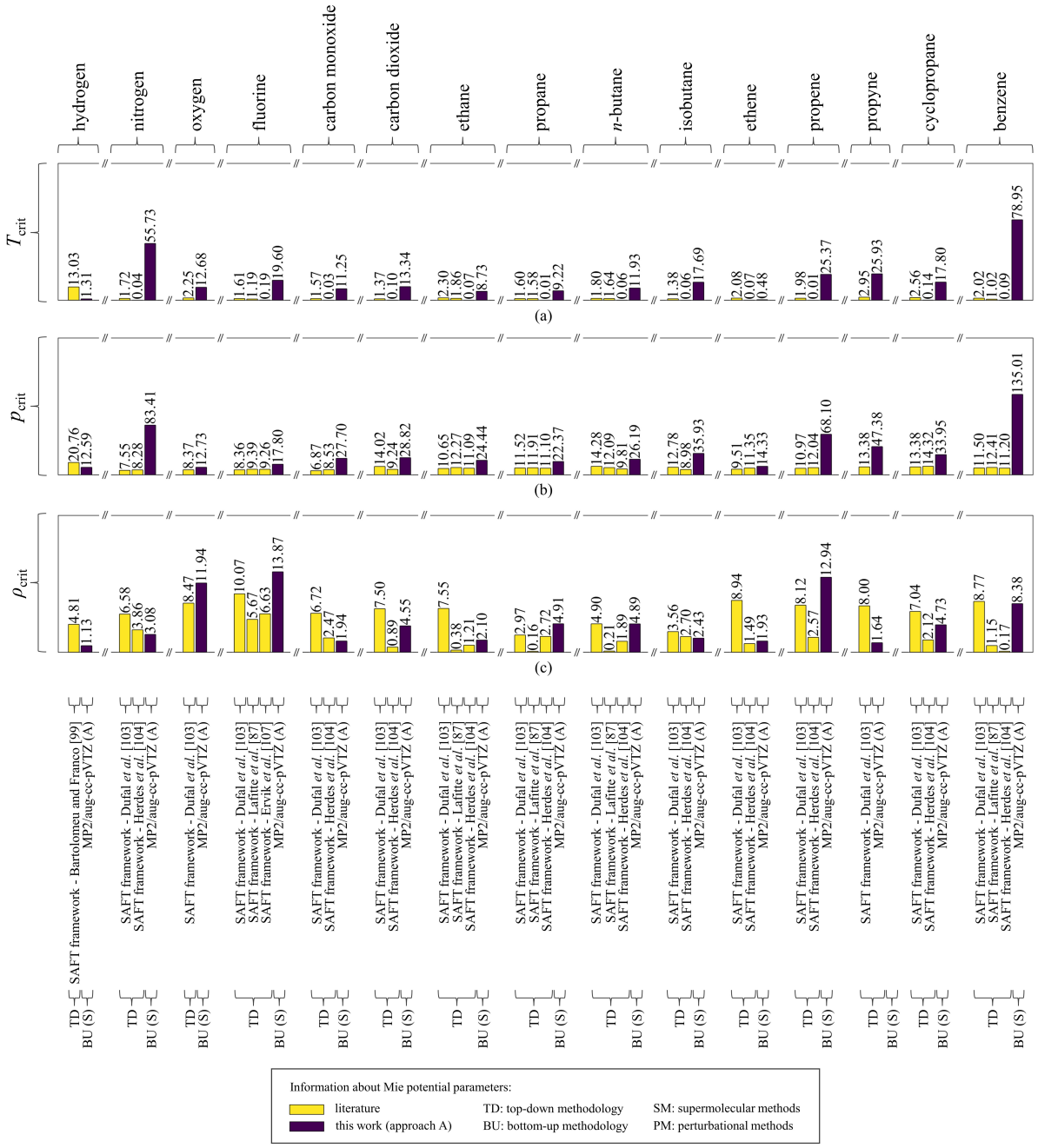

Figure 32 – Percentage deviation of the calculated critical point from NIST reference data using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state for 15 molecules of different chemical classes (hydrogen, nitrogen, oxygen, fluorine, carbon monoxide, carbon dioxide, ethane, propane, *n*-butane, isobutane, ethene, propene, propyne, cyclopropane, and benzene). Results for (a) T_{crit}, (b) p_{crit}, and (c) ρ_{crit}.

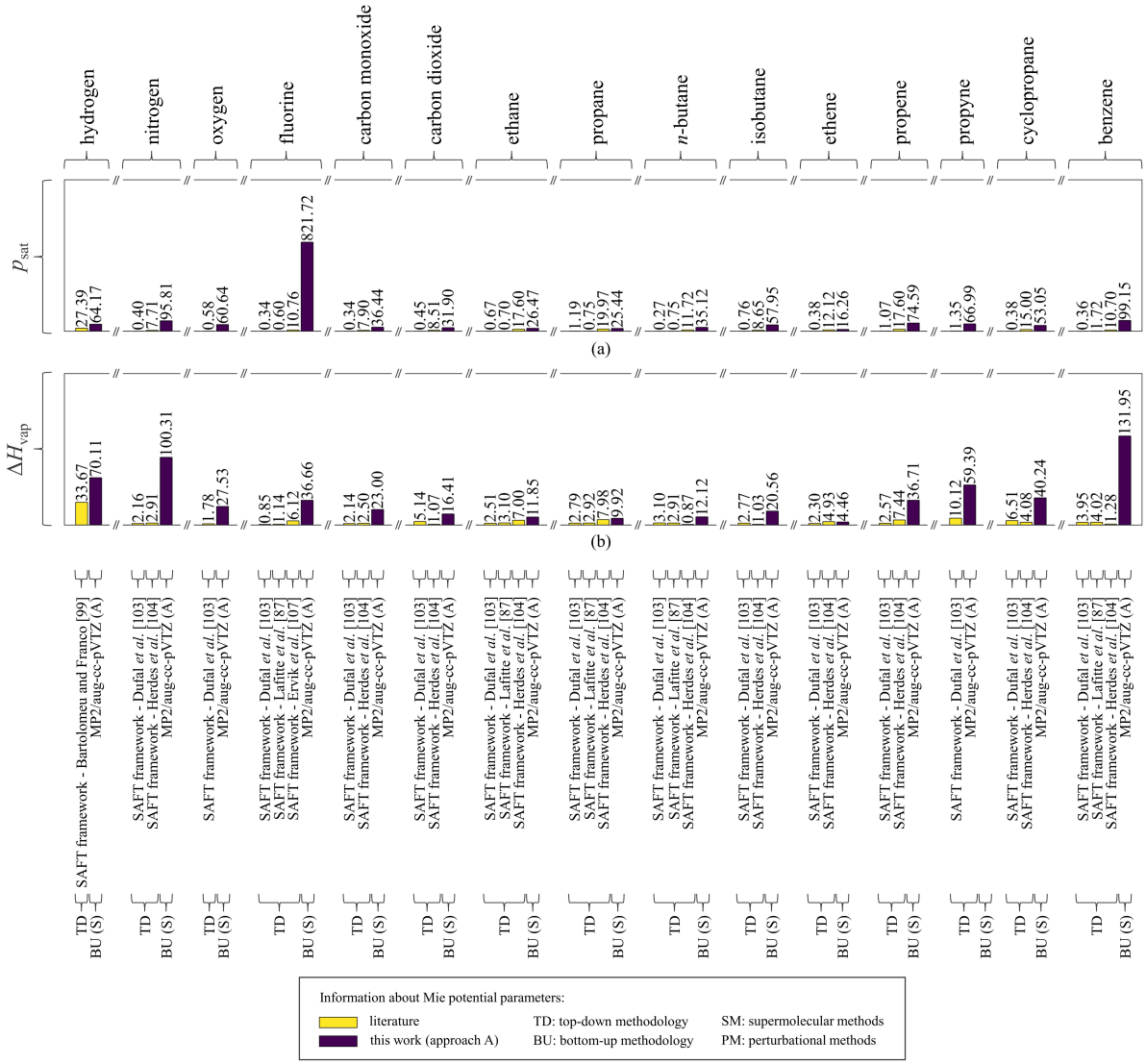

Figure 33 – AARD% values for saturation properties for fifteen molecules of different chemical classes (hydrogen, nitrogen, oxygen, fluorine, carbon monoxide, carbon dioxide, ethane, propane, *n*-butane, isobutane, ethene, propene, propyne, cyclopropane, and benzene) using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state. Results for (a) p_{sat}, (b) ΔH_{vap}.

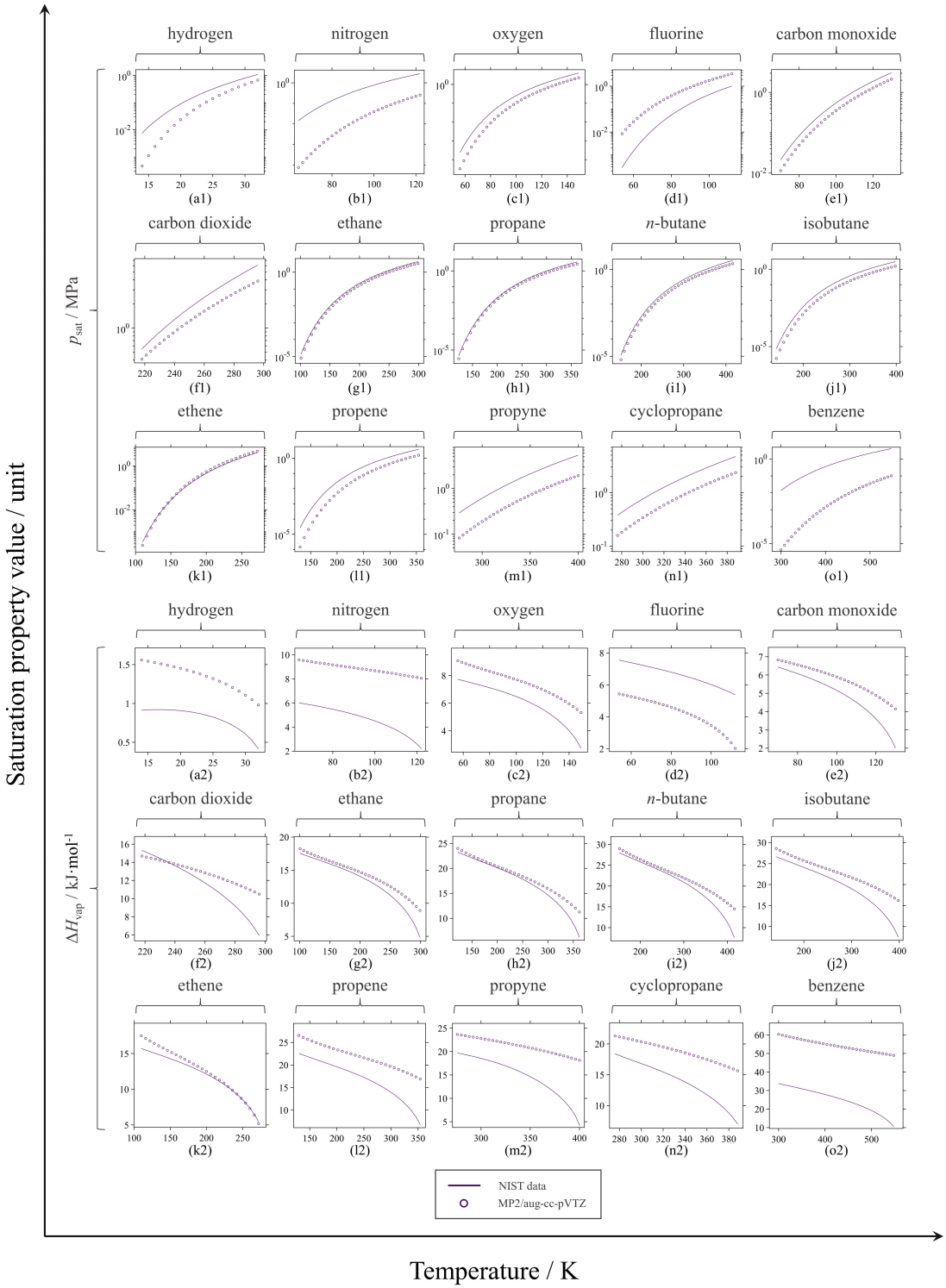

Figure 34 – Predicted saturation properties for 15 molecules of different chemical classes using bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies by approach A in the SAFT-VR Mie equation of state. Results of p_{sat} for (a1) hydrogen, (b1) nitrogen, (c1) oxygen, (d1) fluorine, (e1) carbon monoxide, (f1) carbon dioxide, (g1) ethane, (h1) propane, (i1) *n*-butane, (j1) isobutane, (k1) ethene, (l1) propene, (m1) propyne, (n1) cyclopropane, and (o1) benzene. Results of ΔH_{vap} for (a2) hydrogen, (b2) nitrogen, (c2) oxygen, (d2) fluorine, (e2) carbon monoxide, (f2) carbon dioxide, (g2) ethane, (h2) propane, (i2) *n*-butane, (j2) isobutane, (k2) ethene, (l2) propene, (m2) propyne, (n2) cyclopropane, and (o2) benzene.

3.4 SAFT-VR Mie as pure predictive equation of state

In this Section of this dissertation, the results previously obtained are summarized. In Table 5, the Mie potential parameters are presented for 26 molecules of different chemical classes, which were obtained through calculations *ab initio* to be used in SAFT-VR Mie equation of state and assigned it a purely predictive character. The parameters presented in Table 5 were derived by the approach A of bottom-up methodology from interaction energy curves obtained through MP2/aug-cc-pVTZ theory level for dimers of different spatial configurations, according to the methodology described in Section 2.2.2.

Table 5 – Bottom-up-based Mie potential parameters derived in this study by the approach A from interaction energy curves obtained through MP2/aug-cc-pVTZ theory level for dimers of different spatial configurations for a group of 26 molecules of different chemical classes.

		Mie potential parameters[a]				
Chemical class	Molecule	m	σ/Å	λ_{rep}	λ_{att}	(ε/k_B)/K
light gases	helium	1.0000	2.8475	9.2472	6.0000	6.7890
	neon	1.0000	2.9875	11.4790	6.0000	17.1988
	argon	1.0000	3.4295	13.5842	6.0000	121.8498
	krypton	1.0000	3.6657	13.1544	6.0000	175.0051
	hydrogen	1.0000	3.1995	8.2660	6.0000	21.6242
	nitrogen	1.0000	3.5137	12.3709	6.0000	152.5762
	oxygen	1.0000	3.4915	14.6902	6.0000	148.2514
	fluorine	1.0000	3.3561	13.8871	6.0000	96.0516
	carbon monoxide	1.0000	3.5771	13.4940	6.0000	120.4417
	carbon dioxide	1.0000	3.6965	15.2710	6.0000	298.9670
linear hydrocarbons	methane	1.0000	3.7236	12.1942	6.0000	145.5861
	ethane	1.0000	4.2470	16.4039	6.0000	297.7097
	propane	1.0000	4.7931	19.5709	6.0000	390.0400
	n-butane	1.0000	5.2170	21.3595	6.0000	474.5755
branched hydrocarbons	isobutane	1.0000	5.1661	19.4316	6.0000	462.1039
unsatured hydrocarbons	ethene	1.0000	4.1369	25.9756	6.0000	299.1178
	propene	1.0000	4.3794	18.0501	6.0000	426.8011
	propyne	1.0000	4.4088	16.9949	6.0000	461.6010
cyclic hydrocarbons	cyclopropane	1.0000	4.4508	17.3623	6.0000	431.4277
aromatic hydrocarbons	benzene	1.0000	4.9840	19.8237	6.0000	975.8040
refrigerants	tetrafluormethane	1.0000	4.3246	17.6995	6.0000	218.8569
	fluoromethane	1.0000	4.1530	13.8822	6.0000	259.8926
	chlorotrifluoromethane	1.0000	3.8534	12.0000	6.0000	476.7379
	dichlorodifluoromethane	1.0000	4.9122	18.2487	6.0000	462.8079
	chlorodifluoromethane	1.0000	4.8666	17.5977	6.0000	493.5344
	dichlorofluoromethane	1.0000	4.6220	16.7608	6.0000	344.2269

[a] k_B is the Boltzmann constant

This study demonstrated that the proposed approach A for obtaining the Mie potential parameters through *ab initio* calculations proved to be more adequate. In approach A, a unique set of Mie potential parameters is obtained as an average of the set of parameters fitted to each interaction energy curve of each spatial configuration for each molecule studied. It was observed throughout the investigations perfomed that approach A of the bottom-up methodology generates values of the Mie potential parameters σ and ε close to those obtained by the top-down methodology when fitted directly to reproduce experimental data through the SAFT-VR Mie equation of state.

The bottom-up-based SAFT-VR Mie equation of state parameters obtained in this study were used to predict thermophysical properties of the 26 studied molecules. The predicted values of thermophysical properties such as density (ρ), isochoric heat capacity (c_v), isobaric heat capacity (c_p), speed of sound (c_sound), and Joule-Thomson coefficient (μ_JT) were compared with NIST reference data by computing AARD% values. The NIST reference data considered for calculating AARD% values of thermophysical properties can be consulted in Table C1 found in Appendix C. Fig. 35 summarizes the AARD% values obtained for each studied molecule in predicting thermophysical properties. Alternatively, the information presented in Fig. 35 is also available in Table D13 in Appendix D.

The results presented in Fig. 35 indicate that the methodology developed in this study for obtaining state equation parameters through *ab initio* calculations proved consistent. For the calculation of some thermophysical properties such as ρ, c_v, c_p, and c_sound of methane and substituted-methane compounds, the AARD% values obtained by the SAFT-VR Mie equation of state presented in Fig. 35 were close to those observed through molecular dynamics simulations (previously described in Section 3.2.2). Although for some cases the AARD% values observed in Fig. 35 reach tens of percent and for some specific exceptions reach hundreds of percent, these errors may be associated with the inadequacy of the spherical 1-site coarse-grained model adopted to represent more complex molecules and the inability of the Mie potential to describe purely repulsive interactions.

The capacity of the bottom-up-based SAFT-VR Mie equation of state parameters obtained in this study to predict the critical point was also evaluated. Fig. 36 summarizes the percentage deviations from the critical temperature (T_crit), pressure (p_crit), and density (ρ_crit) relative to NIST reference data for the 26 studied molecules. Alternatively, the information presented in Fig. 36 is also available in Table D14 in Appendix D. In the critical point calculation, the trend observed in percentage deviations remained similar to that observed for the AARD% values of thermophysical properties. For some cases the percentage deviations observed are in tens of percent, with some exceptions also reaching hundreds of percent. This corroborates that some of the impositions considered in the modeling of the investigated molecules may be affecting the observed results (*e.g.*, Mie potential, 1-site coarse-grained spherical model).

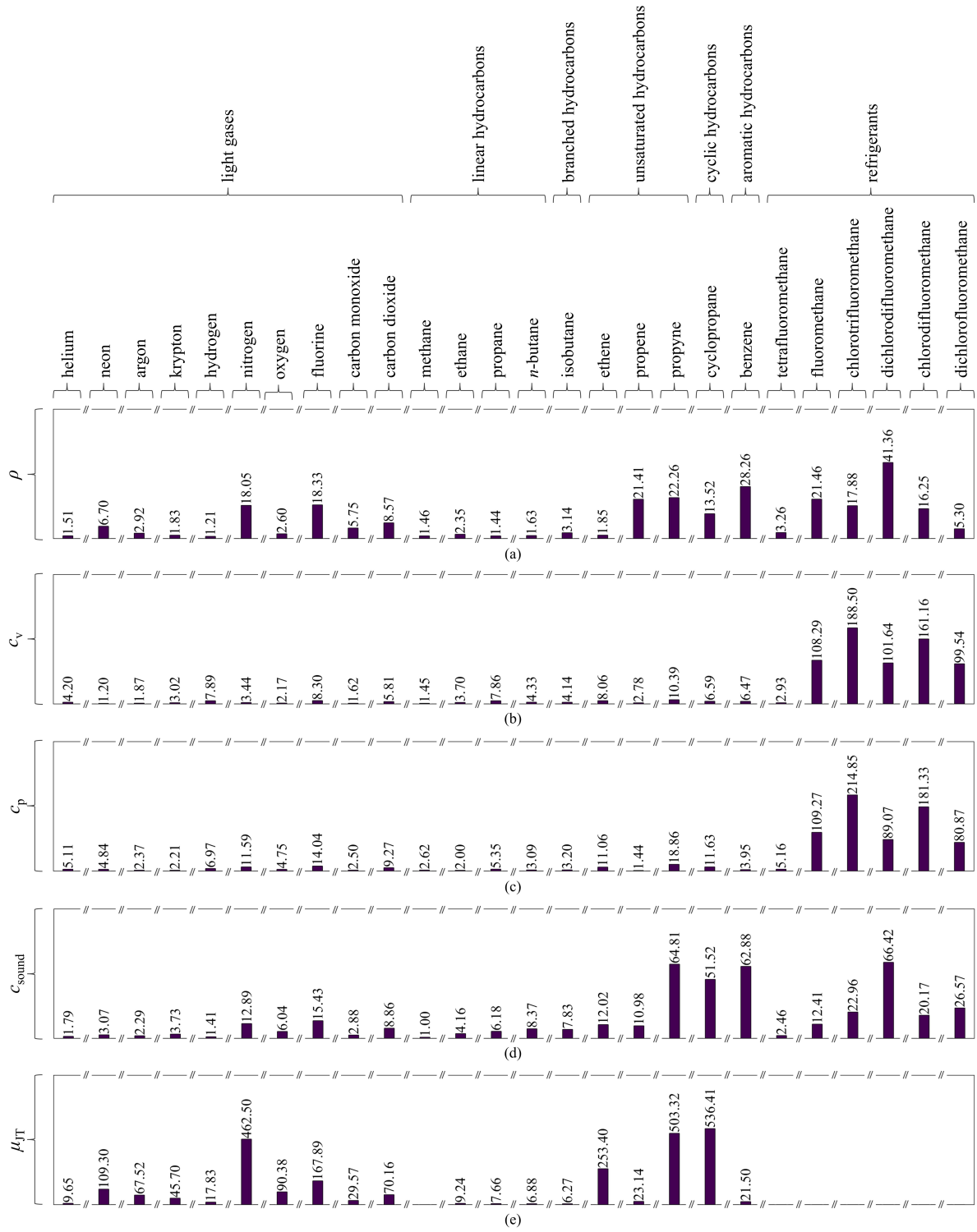

Figure 35 – AARD% values for thermophysical properties of 26 molecules from different chemical classes obtained using a predictive SAFT-VR Mie equation of state with parameters derived from *ab initio* calculations performed in this study. Results of (a) ρ, (b) c_{v}, (c) c_{p}, (d) c_{sound}, and (e) μ_{JT}.

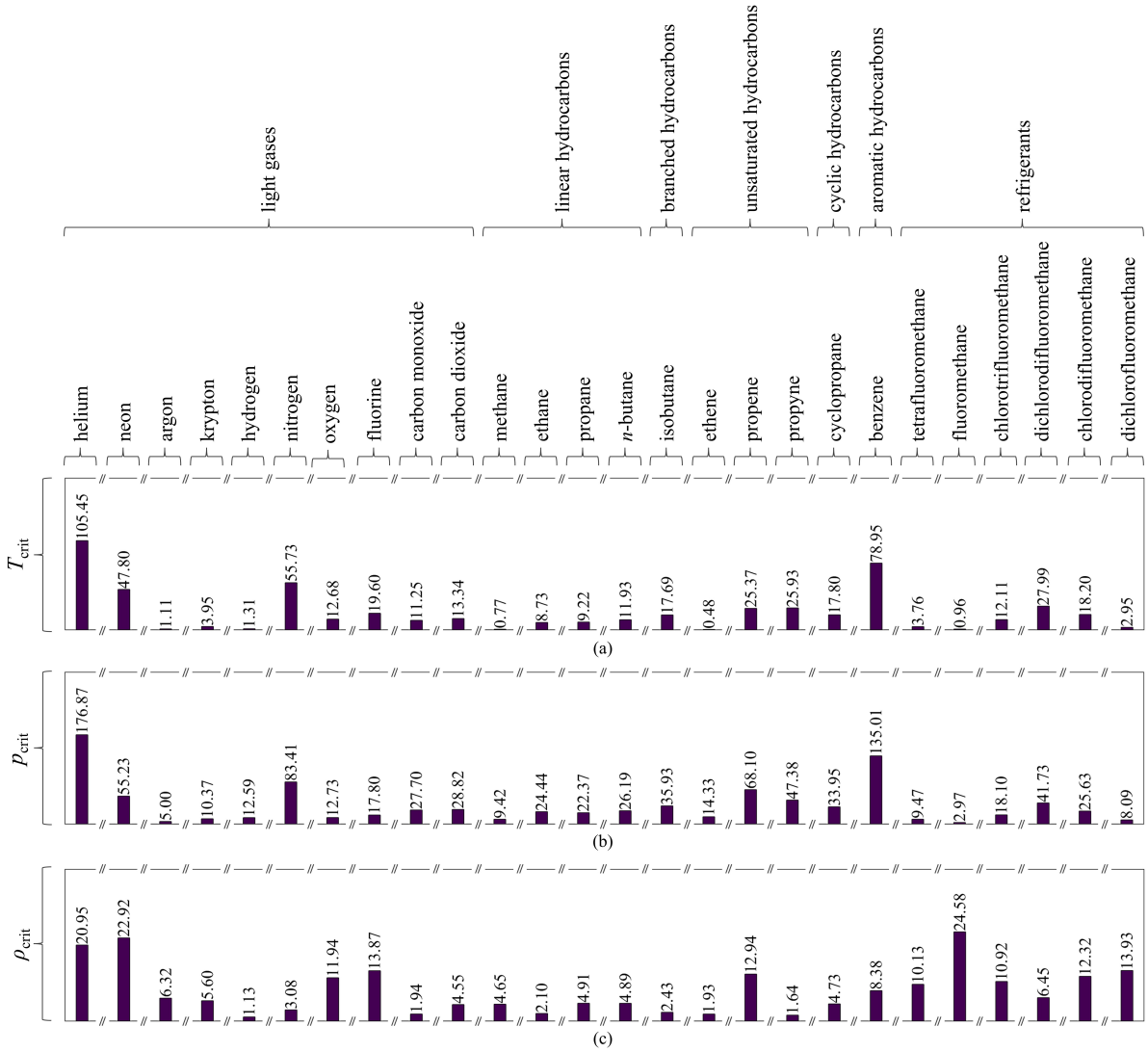

Figure 36 – Percentage deviation in relation to the real critical point of 26 molecules from different chemical classes obtained using a predictive SAFT-VR Mie equation of state with parameters derived from *ab initio* calculations performed in this study. Results for (a) T_{crit}, (b) p_{crit}, and (c) ρ_{crit}.

As a final investigation, the prediction of saturation properties was investigated using the bottom-up-based SAFT-VR Mie equation of state parameters of this study. SAFT-VR Mie equation of state was used to calculate the saturation pressure (p_{sat}) and the enthalpy of vaporization (ΔH_{vap}). The values The NIST reference data considered for calculating AARD% values of saturation properties can be consulted in Table C2 found in Appendix C. The AARD% values referring to saturation properties can be seen in Fig. 37. Alternatively, the information presented in Fig. 37 is also available in Table D15 in Appendix D.

The AARD% values for saturation properties presented in Fig. 37 were, in general, higher than those observed for thermophysical properties. For values referring to p_{sat}, most of the AARD% values obtained are in the order of tens of percent, with exceptions of values with smaller and larger orders of magnitude. It is important to highlight that, for certain temperature ranges, the numerical values of p_{sat} found in NIST are considerably lower than those observed for the other variables referring to thermophysical properties and critical point. This fact combined with the numerical calculation error associated with the prediction of p_{sat} through the SAFT-VR Mie equation of state can significantly impact the AARD% values presented, increasing them. This probable error can be propagated to ΔH_{vap}, since this last property is calculated through p_{sat}.

Although the errors associated with the prediction of thermophysical properties, critical point e saturation properties do not allow immediate use of the purely predictive SAFT-VR Mie equation of state in the design of chemical plants, a major step has been taken in establishing a connection between the quantum and macro scales. At this point, the remaining challenge is to investigate the error associated with the prediction of thermophysical properties, the critical point and saturation properties, which can be further reduced by promoting modifications to the studied model.

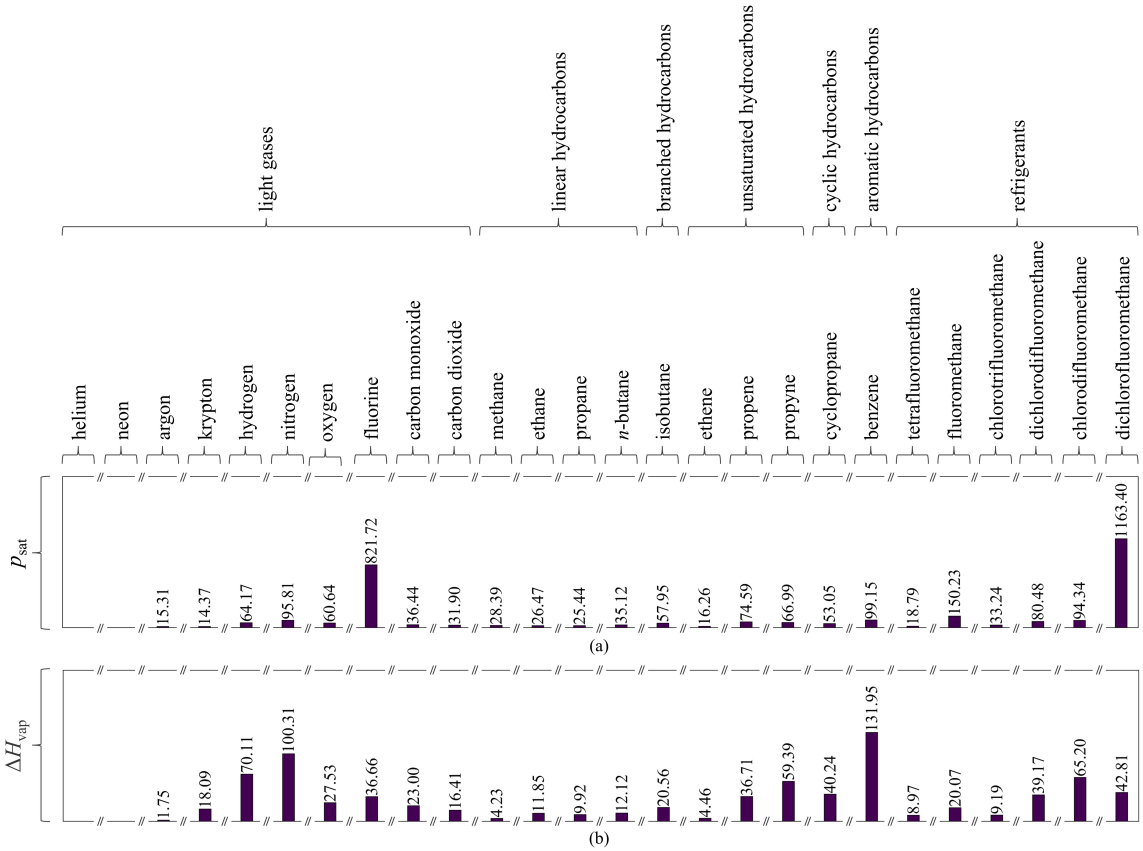

Figure 37 – AARD% values for saturation properties of 26 molecules from different chemical classes obtained using a predictive SAFT-VR Mie equation of state with parameters derived from *ab initio* calculations performed in this study. Results for (a) p_{sat}, (b) ΔH_{vap}.

4 Conclusion and future perspectives

"The important thing is not to stop questioning."

Albert Einstein

This study demonstrated that interatomic potential parameters obtained through *ab initio* calculations usually employed in molecular dynamics simulations can have their application extended as parameters of molecular-based equations of state within the SAFT framework. To achieve this purpose, 26 molecules from different chemical classes were modeled using a spherical 1-site coarse-grained model with interaction based on the Mie potential. These molecules included light gases (helium, neon, argon, and krypton, hydrogen, nitrogen, oxygen, fluorine, carbon monoxide, and carbon dioxide), linear hydrocarbons (methane, ethane, propane, and *n*-butane), branched hydrocarbons (isobutane), unsaturated hydrocarbons (ethene, propene, and propyne), cyclic hydrocarbons (cyclopropane), aromatic hydrocarbons (benzene), and refrigerants (tetrafluoromethane, fluoromethane, chlorotrifluoromethane, dichlorodifluoromethane, dichlorofluoromethane, chlorodifluoromethane). By the bottom-up methodology, the Mie potential parameters for the 26 molecules studied were obtained through the interaction energy curves for dimers of various configurations by *ab initio* calculations via supermolecular and perturbational methods.

Two distinct parameter fitting approaches were considered in the bottom-up methodology for obtaining a unique set of the Mie potential parameters for each molecule studied. In the approach called A, the Mie potential parameters of the molecule represent an average of the fitted parameters for each dimer configuration considered. In the approach called B, the Mie potential parameters are simultaneously fitted to all *ab initio* data of all dimer configurations considered, where the energy of each dimer is weighted according to the Boltzmann energy distribution function. The bottom-up-based Mie potential parameters obtained in this study and the few existing in the literature were compared to top-down-based Mie potential parameters found in the literature within the SAFT framework and directly fitted to represent experimental data.

This study demonstrated that the values of most of the top-down-based Mie potential parameters are very close to those of the approach A bottom-up-based Mie potential parameters. This important observation adds to the literature the understanding of the connection between the quantum scale and the macro scale through a molecular-based equation of state. As a result, what a molecular-based, equation of state within SAFT framework is trying to represent indirectly on the quantum scale is the average interaction energy of dimers of different configurations obtained by *ab initio* calculations.

In addition, the top-down and bottom-up based Mie potential parameters considered in this study were used in both molecular dynamics simulations and the SAFT-VR Mie equation of state to predict thermophysical properties (*i.e.*, density, isochoric heat capacity, isobaric heat capacity, speed of sound, and Joule-Thomson coefficient) and in the SAFT-VR Mie equation of state to predict the critical point, and saturation properties (*i.e.*, saturation pressure and enthalpy of vaporization). The results predicted using top-down and bottom-up-based Mie potential parameters were compared to NIST reference data. For molecules in which the spherical 1-site model adopted has been shown to be suitable, the prediction results obtained when using the approach A bottom-up-based approach Mie potential parameters were comparable to those obtained using top-down-base Mie potential parameters obtained within SAFT framework. This observation suggests that interatomic potential parameters obtained through *ab initio* calculations can also be employed as parameters in molecular-based, equations of state such as SAFT-VR Mie.

As obtaining interaction energy curves through *ab initio* calculations does not directly require experimental data, another result that this study demonstrated is that the use of interatomic potential parameters obtained by the bottom-up methodology can also give to the molecular-based equation of states such as SAFT-VR Mie a purely predictive character. An advantage of molecular-based, equations of state is that the computational cost of solving them is much lower than that of molecular dynamics simulations. Using predictive equations of state can be very interesting from a practical engineering point of view. The dream to be pursued by engineers is to obtain a purely predictive equation of state with parameters obtained through *ab initio* calculations without the need to fit them to experimental data.

Although an important step has been taken in understanding the relationship between top-down and bottom-up methodologies for obtaining interatomic potential parameters to be used in molecular-based, equations of state, the development of studies must continue as there is still room for improvement. Errors on average in the order of tens of percent with specific cases exceeding this magnitude in the calculation of thermophysical properties, critical point, and saturation properties through purely predictive models such as those demonstrated in this dissertation can be large enough to prevent them from being used in the design of chemical plants.

Faced with this remaining challenge, the possibilities regarding improvements and complementary studies are vast. Some of the initial topics that can be investigated include the following items:

- to develop more complex models to represent the set of molecules studied (*e.g.*, to change the interaction potential between molecules or even representing molecules through more than one type of site, as in SAFT-γ Mie equation of state in which heteronuclear segments can form each molecule)

- to investigate the influence of quantum effects to construct an effective potential using the Feynman–Hibbs correction as proposed by Aasen *et al* [108].
- to extract other information from *ab initio* calculations in addition to dimer interaction energy curves to be used to obtain bottom-up-based interatomic potential parameters (*e.g.*, to fit the volume of the molecule obtained by calculations *ab initio* in order to optimize the number of segments in the SAFT-VR Mie equation of state).
- to investigate the application of bottom-up-based interatomic potential parameters in calculating mixture properties (*e.g.*, to verify whether the bottom-up-based interatomic potential parameters obtained for mixtures follow any specific combination rule)

Bibliography

1 NEZBEDA, I. On molecular-based equations of state: Perturbation theories, simple models, and saft modeling. *Frontiers in Physics*, v. 8, 2020.

2 CHAPMAN, W. G.; GUBBINS, K. E.; JACKSON, G.; RADOSZ, M. Saft: Equation-of-state solution model for associating fluids. *Fluid Phase Equilibria*, v. 52, p. 31–38, 1989.

3 CHAPMAN, W. G.; GUBBINS, K. E.; JACKSON, G.; RADOSZ, M. New reference equation of state for associating liquids. *Industrial & Engineering Chemistry Research*, v. 29, n. 8, p. 1709–1721, 1990.

4 VAN NHU, N.; SINGH, M.; LEONHARD, K. Quantum mechanically based estimation of perturbed-chain polar statistical associating fluid theory parameters for analyzing their physical significance and predicting properties. *The Journal of Physical Chemistry B*, v. 112, n. 18, p. 5693–5701, 2008.

5 CASSENS, J.; RUETHER, F.; LEONHARD, K.; SADOWSKI, G. Solubility calculation of pharmaceutical compounds – a priori parameter estimation using quantum-chemistry. *Fluid Phase Equilibria*, v. 299, n. 1, p. 161–170, 2010.

6 VON MÜLLER, A.; LEONHARD, K. Surface tension calculations by means of a pcp-saft-dft formalism using equation of state parameters from quantum mechanics. *Fluid Phase Equilibria*, v. 356, p. 96–101, 2013.

7 UMER, M.; ALBERS, K.; SADOWSKI, G.; LEONHARD, K. Pc-saft parameters from ab initio calculations. *Fluid Phase Equilibria*, v. 362, p. 41–50, 2014.

8 KAMINSKI, S.; BARDOW, A.; LEONHARD, K. The trade-off between experimental effort and accuracy for determination of pcp-saft parameters. *Fluid Phase Equilibria*, v. 428, p. 182–189, 2016.

9 KAMINSKI, S.; LEONHARD, K. Sepp: Segment-based equation of state parameter prediction. *Journal of Chemical & Engineering Data*, v. 65, n. 12, p. 5830–5843, 2020.

10 WALKER, P. J.; ZHAO, T.; HASLAM, A. J.; JACKSON, G. Ab initio development of generalized lennard-jones (mie) force fields for predictions of thermodynamic properties in advanced molecular-based saft equations of state. *The Journal of Chemical Physics*, v. 156, n. 15, p. 154106, 2022.

11 GUBBINS, K. E.; MOORE, J. D. Molecular modeling of matter: Impact and prospects in engineering. *Industrial & Engineering Chemistry Research*, v. 49, n. 7, p. 3026–3046, 2010.

12 PALMER, J. C.; DEBENEDETTI, P. G. Recent advances in molecular simulation: A chemical engineering perspective. *AIChE Journal*, v. 61, n. 2, p. 370–383, 2015.

13 MAGINN, E. J. From discovery to data: What must happen for molecular simulation to become a mainstream chemical engineering tool. *AIChE Journal*, v. 55, n. 6, p. 1304–1310, 2009.

14 TUCKERMAN, M. E.; MARTYNA, G. J. Understanding modern molecular dynamics: Techniques and applications. *The Journal of Physical Chemistry B*, v. 104, n. 2, p. 159–178, 2000.

15 MAGINN, E. J.; ELLIOTT, J. R. Historical perspective and current outlook for molecular dynamics as a chemical engineering tool. *Industrial & Engineering Chemistry Research*, v. 49, n. 7, p. 3059–3078, 2010.

16 IFTIMIE, R.; MINARY, P.; TUCKERMAN, M. E. Ab initio molecular dynamics: Concepts, recent developments, and future trends. *Proceedings of the National Academy of Sciences*, v. 102, n. 19, p. 6654–6659, 2005.

17 THEODOROU, D. N. Progress and outlook in monte carlo simulations. *Industrial & Engineering Chemistry Research*, v. 49, n. 7, p. 3047–3058, 2010.

18 BROOKS, C. L.; CASE, D. A.; PLIMPTON, S.; ROUX, B.; SPOEL, D. van der; TAJKHORSHID, E. Classical molecular dynamics. *The Journal of Chemical Physics*, v. 154, n. 10, p. 100401, 2021.

19 SCHRÖDINGER, E. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, American Physical Society, v. 28, p. 1049–1070, 1926.

20 HARTREE, D. R. The wave mechanics of an atom with a non-coulomb central field. part ii. some results and discussion. *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, v. 24, n. 1, p. 111–132, 1928.

21 SLATER, J. C. The self consistent field and the structure of atoms. *Physical Review*, American Physical Society, v. 32, p. 339–348, 1928.

22 SLATER, J. C. Note on hartree's method. *Physical Review*, American Physical Society, v. 35, p. 210–211, 1930.

23 FOCK, V. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, v. 61, p. 126–148, 1930.

24 FOCK, V. elfconsistent field“ mit austausch für natrium. *Zeitschrift für Physik*, v. 62, p. 795–805, 1930.

25 HOHENBERG, P.; KOHN, W. Inhomogeneous electron gas. *Physical Review*, American Physical Society, v. 136, p. B864–B871, 1964.

26 KOHN, W.; SHAM, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review*, American Physical Society, v. 140, p. A1133–A1138, 1965.

27 GONZÁLEZ, M.A. Force fields and molecular dynamics simulations. *JDN*, v. 12, p. 169–200, 2011.

28 JEZIORSKI, B.; MOSZYNSKI, R.; SZALEWICZ, K. Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes. *Chemical Reviews*, v. 94, n. 7, p. 1887–1930, 1994.

29 HOHENSTEIN, E. G.; SHERRILL, C. D. Wavefunction methods for noncovalent interactions. *WIREs Computational Molecular Science*, v. 2, n. 2, p. 304–326, 2012.

30 PARKER, T. M.; BURNS, L. A.; PARRISH, R. M.; RYNO, A. G.; SHERRILL, C. D. Levels of symmetry adapted perturbation theory (sapt). i. efficiency and performance for interaction energies. *The Journal of Chemical Physics*, v. 140, n. 9, p. 094106, 2014.

31 GRIMME, S. Accurate description of van der waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry*, v. 25, n. 12, p. 1463–1473, 2004.

32 GRIMME, S. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, v. 27, n. 15, p. 1787–1799, 2006.

33 GRIMME, S.; ANTONY, J.; EHRLICH, S.; KRIEG, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, v. 132, n. 15, p. 154104, 2010.

34 GRIMME, S. Improved second-order møller–plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *The Journal of Chemical Physics*, v. 118, n. 20, p. 9095–9102, 2003.

35 MARCHETTI, O.; WERNER, H.-J. Accurate calculations of intermolecular interaction energies using explicitly correlated coupled cluster wave functions and a dispersion-weighted mp2 method. *The Journal of Physical Chemistry A*, v. 113, n. 43, p. 11580–11585, 2009.

36 PURVIS, G. D.; BARTLETT, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *The Journal of Chemical Physics*, v. 76, n. 4, p. 1910–1918, 1982.

37 RAGHAVACHARI, K.; TRUCKS, G. W.; POPLE, J. A.; HEAD-GORDON, M. A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, v. 157, n. 6, p. 479–483, 1989.

38 JOSHI, S. Y.; DESHMUKH, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Molecular Simulation*, Taylor & Francis, v. 47, n. 10-11, p. 786–803, 2021.

39 WERTHEIM, M. . S. Fluids with highly directional attractive forces. i. statistical thermodynamics. *Journal of Statistical Physics*, v. 35, n. 1, p. 19–34, 1984.

40 WERTHEIM, M. S. Fluids with highly directional attractive forces. ii. thermodynamic perturbation theory and integral equations. *Journal of Statistical Physics*, v. 35, n. 1, p. 35–47, 1984.

41 WERTHEIM, M. S. Fluids with highly directional attractive forces. iii. multiple attraction sites. *Journal of Statistical Physics*, v. 42, n. 3, p. 459–476, 1986.

42 WERTHEIM, M. S. Fluids with highly directional attractive forces. iv. equilibrium polymerization. *Journal of Statistical Physics*, v. 42, n. 3, p. 477–492, 1986.

43 WERTHEIM, M. S. Thermodynamic perturbation theory of polymerization. *The Journal of Chemical Physics*, v. 87, n. 12, p. 7323–7331, 1987.

44 MÜLLER, E. A.; GUBBINS, K. E. Molecular-based equations of state for associating fluids: A review of saft and related approaches. *Industrial & Engineering Chemistry Research*, v. 40, n. 10, p. 2193–2211, 2001.

45 SHAAHMADI, F.; SMITH, S. A.; SCHWARZ, C. E.; BURGER, A. J.; CRIPWELL, J. T. Group-contribution saft equations of state: A review. *Fluid Phase Equilibria*, v. 565, p. 113674, 2023.

46 LYRA, E. P.; FRANCO, L. F. M. Deriving force fields with a multiscale approach: From ab initio calculations to molecular-based equations of state. *The Journal of Chemical Physics*, v. 157, n. 11, p. 114107, 2022.

47 ROOTHAAN, C. C. J. New developments in molecular orbital theory. *Reviews of Modern Physics*, American Physical Society, v. 23, p. 69–89, 1951.

48 HALL, G. G.; LENNARD-JONES, J. E. The molecular orbital theory of chemical valency viii. a method of calculating ionization potentials. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, v. 205, n. 1083, p. 541–552, 1951.

49 MØLLER, C.; PLESSET, M. S. Note on an approximation treatment for many-electron systems. *Physical Review*, American Physical Society, v. 46, p. 618–622, 1934.

50 ČÍŽEK, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *The Journal of Chemical Physics*, v. 45, n. 11, p. 4256–4266, 1966.

51 PALDUS, J.; ČÍŽEK, J.; SHAVITT, I. Correlation problems in atomic and molecular systems. iv. extended coupled-pair many-electron theory and its application to the bh_3 molecule. *Physical Review A*, American Physical Society, v. 5, p. 50–67, 1972.

52 THOMAS, L. H. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, v. 23, n. 5, p. 542–548, 1927.

53 FERMI, E. Un metodo statistico per la determinazione di alcune priorieta dell'atome. *Rend. Accad. Naz. Lincei*, v. 6, n. 602-607, p. 32, 1927.

54 DIRAC, P. A. M. Note on exchange phenomena in the thomas atom. *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, v. 26, n. 3, p. 376–385, 1930.

55 PATKOWSKI, K. Recent developments in symmetry-adapted perturbation theory. *WIREs Computational Molecular Science*, v. 10, n. 3, p. e1452, 2020.

56 DAVIDSON, E. R.; FELLER, D. Basis set selection for molecular calculations. *Chemical Reviews*, v. 86, n. 4, p. 681–696, 1986.

57 JENSEN, F. Atomic orbital basis sets. *WIREs Computational Molecular Science*, v. 3, n. 3, p. 273–295, 2013.

58 DUNNING THOM H., J. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics*, v. 90, n. 2, p. 1007–1023, 1989.

59 GOLD, V.; BETHELL, D. Calculation of molecular structure and energy by force-field methods. In: . [S.l.]: Academic Press, 1976, (Advances in Physical Organic Chemistry, v. 13). p. 1–82.

60 ALLINGER, N. L. Conformational analysis. 130. mm2. a hydrocarbon force field utilizing v1 and v2 torsional terms. *Journal of the American Chemical Society*, v. 99, n. 25, p. 8127–8134, 1977.

61 LII, J. H.; ALLINGER, N. L. Molecular mechanics. the mm3 force field for hydrocarbons. 2. vibrational frequencies and thermodynamics. *Journal of the American Chemical Society*, v. 111, n. 23, p. 8566–8575, 1989.

62 ALLINGER, N. L.; YUH, Y. H.; LII, J. H. Molecular mechanics. the mm3 force field for hydrocarbons. 1. *Journal of the American Chemical Society*, v. 111, n. 23, p. 8551–8566, 1989.

63 NEVINS, N.; CHEN, K.; ALLINGER, N. L. Molecular mechanics (mm4) calculations on alkenes. *Journal of Computational Chemistry*, v. 17, n. 5-6, p. 669–694, 1996.

64 ALLINGER, N. L.; CHEN, K.; LII, J.-H. An improved force field (mm4) for saturated hydrocarbons. *Journal of Computational Chemistry*, v. 17, n. 5-6, p. 642–668, 1996.

65 ALLINGER, N. L.; CHEN, K.; KATZENELLENBOGEN, J. A.; WILSON, S. R.; ANSTEAD, G. M. Hyperconjugative effects on carbon—carbon bond lengths in molecular mechanics (mm4). *Journal of Computational Chemistry*, v. 17, n. 5-6, p. 747–755, 1996.

66 ALLINGER, N. L.; CHEN, K.-H.; LII, J.-H.; DURKIN, K. A. Alcohols, ethers, carbohydrates, and related compounds. i. the mm4 force field for simple compounds. *Journal of Computational Chemistry*, v. 24, n. 12, p. 1447–1472, 2003.

67 MAYO, S. L.; OLAFSON, B. D.; GODDARD, W. A. Dreiding: a generic force field for molecular simulations. *The Journal of Physical Chemistry*, v. 94, n. 26, p. 8897–8909, 1990.

68 RAPPE, A. K.; CASEWIT, C. J.; COLWELL, K. S.; GODDARD, W. A. I.; SKIFF, W. M. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, v. 114, n. 25, p. 10024–10035, 1992.

69 CASEWIT, C. J.; COLWELL, K. S.; RAPPE, A. K. Application of a universal force field to organic molecules. *Journal of the American Chemical Society*, v. 114, n. 25, p. 10035–10046, 1992.

70 BROOKS, B. R.; BRUCCOLERI, R. E.; OLAFSON, B. D.; STATES, D. J.; SWAMINATHAN, S.; KARPLUS, M. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, v. 4, n. 2, p. 187–217, 1983.

71 CORNELL, W. D.; CIEPLAK, P.; BAYLY, C. I.; GOULD, I. R.; MERZ, K. M.; FERGUSON, D. M.; SPELLMEYER, D. C.; FOX, T.; CALDWELL, J. W.; KOLLMAN, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, v. 117, n. 19, p. 5179–5197, 1995.

72 SCOTT, W. R. P.; HüNENBERGER, P. H.; TIRONI, I. G.; MARK, A. E.; BILLETER, S. R.; FENNEN, J.; TORDA, A. E.; HUBER, T.; KRüGER, P.; GUNSTEREN, W. F. van. The gromos biomolecular simulation program package. *The Journal of Physical Chemistry A*, v. 103, n. 19, p. 3596–3607, 1999.

73 JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, v. 118, n. 45, p. 11225–11236, 1996.

74 SUN, H. Compass: An ab initio force-field optimized for condensed-phase applicationsoverview with details on alkane and benzene compounds. *The Journal of Physical Chemistry B*, v. 102, n. 38, p. 7338–7364, 1998.

75 VERLET, L. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical Review*, American Physical Society, v. 159, p. 98–103, 1967.

76 VERLET, L. Computer "experiments" on classical fluids. ii. equilibrium correlation functions. *Physical Review.*, American Physical Society, v. 165, p. 201–214, 1968.

77 SWOPE, W. C.; ANDERSEN, H. C.; BERENS, P. H.; WILSON, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, v. 76, n. 1, p. 637–649, 1982.

78 GUNSTEREN, W. F. V.; BERENDSEN, H. J. C. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, Taylor & Francis, v. 1, n. 3, p. 173–185, 1988.

79 HUANG, S. H.; RADOSZ, M. Equation of state for small, large, polydisperse, and associating molecules. *Industrial & Engineering Chemistry Research*, v. 29, n. 11, p. 2284–2294, 1990.

80 HUANG, S. H.; RADOSZ, M. Equation of state for small, large, polydisperse, and associating molecules: extension to fluid mixtures. *Industrial & Engineering Chemistry Research*, v. 30, n. 8, p. 1994–2005, 1991.

81 GIL-VILLEGAS, A.; GALINDO, A.; WHITEHEAD, P. J.; MILLS, S. J.; JACKSON, G.; BURGESS, A. N. Statistical associating fluid theory for chain molecules with attractive potentials of variable range. *The Journal of Chemical Physics*, v. 106, n. 10, p. 4168–4186, 1997.

82 GROSS, J.; SADOWSKI, G. Application of perturbation theory to a hard-chain reference fluid: an equation of state for square-well chains. *Fluid Phase Equilibria*, v. 168, n. 2, p. 183–199, 2000.

83 GROSS, J.; SADOWSKI, G. Perturbed-chain saft: An equation of state based on a perturbation theory for chain molecules. *Industrial & Engineering Chemistry Research*, v. 40, n. 4, p. 1244–1260, 2001.

84 GROSS, J.; SADOWSKI, G. Modeling polymer systems using the perturbed-chain statistical associating fluid theory equation of state. *Industrial & Engineering Chemistry Research*, v. 41, p. 1084–1093, 2002.

85 LAFITTE, T.; BESSIERES, D.; PIÑEIRO, M. M.; DARIDON, J.-L. Simultaneous estimation of phase behavior and second-derivative properties using the statistical associating fluid theory with variable range approach. *The Journal of Chemical Physics*, v. 124, n. 2, p. 024509, 2006.

86 LAFITTE, T.; PIÑEIRO, M. M.; DARIDON, J.-L.; BESSIÈRES, D. A comprehensive description of chemical association effects on second derivative properties of alcohols through a saft-vr approach. *The Journal of Physical Chemistry B*, v. 111, n. 13, p. 3447–3461, 2007.

87 LAFITTE, T.; APOSTOLAKOU, A.; AVENDAÑO, C.; GALINDO, A.; ADJIMAN, C. S.; MÜLLER, E. A.; JACKSON, G. Accurate statistical associating fluid theory for chain molecules formed from mie segments. *The Journal of Chemical Physics*, v. 139, n. 15, p. 154504, 2013.

88 PAPAIOANNOU, V.; LAFITTE, T.; AVENDAÑO, C.; ADJIMAN, C. S.; JACKSON, G.; MÜLLER, E. A.; GALINDO, A. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from mie segments. *The Journal of Chemical Physics*, v. 140, n. 5, p. 054107, 2014.

89 CHEN, S. S.; KREGLEWSKI, A. Applications of the augmented van der waals theory of fluids.: I. pure fluids. *Berichte der Bunsengesellschaft für physikalische Chemie*, v. 81, n. 10, p. 1048–1052, 1977.

90 FRISCH, M. J.; TRUCKS, G. W.; SCHLEGEL, H. B.; SCUSERIA, G. E.; ROBB, M. A.; CHEESEMAN, J. R.; SCALMANI, G.; BARONE, V.; MENNUCCI, B.; PETERSSON, G. A.; NAKATSUJI, H.; CARICATO, M.; LI, X.; HRATCHIAN, H. P.; IZMAYLOV, A. F.; BLOINO, J.; ZHENG, G.; SONNENBERG, J. L.; HADA, M.; EHARA, M.; TOYOTA, K.; FUKUDA, R.; HASEGAWA, J.; ISHIDA, M.; NAKAJIMA, T.; HONDA, Y.; KITAO, O.; NAKAI, H.; VREVEN, T.; MONTGOMERY Jr., J. A.; PERALTA, J. E.; OGLIARO, F.; BEARPARK, M.; HEYD, J. J.; BROTHERS, E.; KUDIN, K. N.; STAROVEROV, V. N.; KOBAYASHI, R.; NORMAND, J.; RAGHAVACHARI, K.; RENDELL, A.; BURANT, J. C.; IYENGAR, S. S.; TOMASI, J.; COSSI, M.; REGA, N.; MILLAM, J. M.; KLENE, M.; KNOX, J. E.; CROSS, J. B.; BAKKEN, V.; ADAMO, C.; JARAMILLO, J.; GOMPERTS, R.; STRATMANN, R. E.; YAZYEV, O.; AUSTIN, A. J.; CAMMI, R.; POMELLI, C.; OCHTERSKI, J. W.; MARTIN, R. L.; MOROKUMA, K.; ZAKRZEWSKI, V. G.; VOTH, G. A.; SALVADOR, P.; DANNENBERG, J. J.; DAPPRICH, S.; DANIELS, A. D.; FARKAS, O.; FORESMAN, J. B.; ORTIZ, J. V.; CIOSLOWSKI, J.; FOX, D. J. *Gaussian 09 Revision E.01*. Wallingford, CT, 2009.

91 SMITH, D. G. A.; BURNS, L. A.; SIMMONETT, A. C.; PARRISH, R. M.; SCHIEBER, M. C.; GALVELIS, R.; KRAUS, P.; KRUSE, H.; REMIGIO, R. D.; ALENAIZAN, A.; JAMES, A. M.; LEHTOLA, S.; MISIEWICZ, J. P.; SCHEURER, M.; SHAW, R. A.; SCHRIBER, J. B.; XIE, Y.; GLICK, Z. L.; SIRIANNI, D. A.; O'BRIEN, J. S.; WALDROP, J. M.; KUMAR, A.; HOHENSTEIN, E. G.; PRITCHARD, B. P.; BROOKS, B. R.; SCHAEFER, H. F.; SOKOLOV, A. Y.; PATKOWSKI, K.; DEPRINCE, A. E.; BOZKAYA, U.; KING, R. A.; EVANGELISTA, F. A.; TURNEY, J. M.; CRAWFORD, T. D.; SHERRILL, C. D. Psi4 1.4: Open-source software for high-throughput quantum chemistry. *The Journal of Chemical Physics*, v. 152, n. 18, p. 184108, 2020.

92 CHAO, S.-W.; LI, A. H.-T.; CHAO, S. D. Molecular dynamics simulations of fluid methane properties using ab initio intermolecular interaction potentials. *Journal of Computational Chemistry*, v. 30, n. 12, p. 1839–1849, 2009.

93 LEVENBERG, K. *Quarterly of Applied Mathematics*, v. 2, n. 2, p. 164–168, 1944.

94 MARQUARDT, D. W. *Journal of the Society for Industrial and Applied Mathematics*, v. 11, n. 2, p. 431–441, 1963.

95 LEMMON, E. W.; BELL, I.; HUBER, M. L.; MCLINDEN, M. O. *NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 10.0, National Institute of Standards and Technology*. 2018.

96 THOMPSON, A. P.; AKTULGA, H. M.; BERGER, R.; BOLINTINEANU, D. S.; BROWN, W. M.; CROZIER, P. S.; in 't Veld, P. J.; KOHLMEYER, A.; MOORE, S. G.; NGUYEN, T. D.; SHAN, R.; STEVENS, M. J.; TRANCHIDA, J.; TROTT, C.; PLIMPTON, S. J. Lammps - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, v. 271, p. 108171, 2022.

97 JEWETT, A. I.; STELTER, D.; LAMBERT, J.; SALADI, S. M.; ROSCIONI, O. M.; RICCI, M.; AUTIN, L.; MARITAN, M.; BASHUSQEH, S. M.; KEYES, T.; DAME, R. T.; SHEA, J.-E.; JENSEN, G. J.; GOODSELL, D. S. Moltemplate: A tool for coarse-grained modeling of complex biological matter and soft condensed matter physics. *Journal of Molecular Biology*, v. 433, n. 11, p. 166841, 2021.

98 MARTÍNEZ, L.; ANDRADE, R.; BIRGIN, E. G.; MARTíNEZ, J. M. Packmol: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, v. 30, n. 13, p. 2157–2164, 2009.

99 BARTOLOMEU, R. A.; FRANCO, L. F. Thermophysical properties of supercritical h_2 from molecular dynamics simulations. *International Journal of Hydrogen Energy*, v. 45, n. 33, p. 16372–16380, 2020.

100 POLING, B. E.; PRAUSNITZ, J. M.; O'CONNELL, J. P. *The properties of gases and liquids*. 5th. ed. New York: McGraw-Hill, 2001.

101 ALLEN, M. P.; J., T. D. *Computer simulation of liquids*. 2nd. ed. Oxford: Oxford University Press, 2017.

102 LLOVELL, F.; VEGA, L. F. Prediction of thermodynamic derivative properties of pure fluids through the soft-saft equation of state. *The Journal of Physical Chemistry B*, v. 110, n. 23, p. 11427–11437, 2006.

103 DUFAL, S.; LAFITTE, T.; GALINDO, A.; JACKSON, G.; HASLAM, A. J. Developing intermolecular-potential models for use with the saft-vr mie equation of state. *AIChE Journal*, v. 61, n. 9, p. 2891–2912, 2015.

104 HERDES, C.; TOTTON, T. S.; MÜLLER, E. A. Coarse grained force field for the molecular simulation of natural gases and condensates. *Fluid Phase Equilibria*, v. 406, p. 91–100, 2015.

105 HIROTA, E. Anharmonic potential function and equilibrium structure of methane. *Journal of Molecular Spectroscopy*, v. 77, n. 2, p. 213–221, 1979.

106 MEJÍA, A.; HERDES, C.; MÜLLER, E. A. Force fields for coarse-grained molecular simulations from a corresponding states correlation. *Industrial & Engineering Chemistry Research*, v. 53, n. 10, p. 4131–4141, 2014.

107 ERVIK, Å.; MEJÍA, A.; MÜLLER, E. A. Bottled saft: A web app providing saft-γ mie force field parameters for thousands of molecular fluids. *Journal of Chemical Information and Modeling*, v. 56, n. 9, p. 1609–1614, 2016.

108 AASEN, A.; HAMMER, M.; ERVIK, Å.; MÜLLER, E. A.; WILHELMSEN, Ø. Equation of state and force fields for Feynman–Hibbs-corrected Mie fluids. I. Application to pure helium, neon, hydrogen, and deuterium. *The Journal of Chemical Physics*, v. 151, n. 6, p. 064508, 2019.

Appendix

APPENDIX A

Example scripts for *ab initio* calculations

Description: This Appendix briefly presents example scripts for *ab initio* calculations. On the following pages of this Appendix, example scripts are presented used for methane to optimize the geometry and to obtain interaction energies through software for *ab initio* calculations such as Gaussian 09 [90] and psi4 [91].

```
#p scf=tight opt=tight mp2/aug-cc-pvqz

 Job Title

0 1
C
H   1 B1
H   1 B2 2 A2
H   1 B3 2 A3 3 D3
H   1 B4 2 A4 3 D4
Variables:
B1         1.09000
B2         1.09000
A2       109.47122
B3         1.09000
A3       109.47122
D3       240.00000
B4         1.09000
A4       109.47122
D4       120.00000
```

Source code 1 – Example of script (.com file extension) used to optimize methane geometry (at MP2/aug-cc-pVQZ theory level) using Gaussian 09 software [90].

```
molecule ch4_dimer {
   C 0.00000000 0.00000000 0.00000000
   H 0.00000000 1.08452710 0.00000000
   H -0.51125100 -0.36150900 0.88551270
   H -0.51125100 -0.36150900 -0.88551270
   H 1.02250190 -0.36150900 0.00000000
   --
   C 0.00000000 3.77900000 -0.00000000
   H 0.00000000 2.69447290 -0.00000000
   H -0.51125100 4.14050900 -0.88551270
   H -0.51125100 4.14050900 0.88551270
   H 1.02250190 4.14050900 0.00000000
}

set basis aug-cc-pVTZ

E = energy('MP2',molecule=ch4_dimer, bsse_type='cp')
Efinal = E* psi_hartree2kcalmol
psi4.print_out("MP2/aug-cc-pVTZ = ")
psi4.print_out("%10.6f" % (Efinal))
```

Source code 2 – Example of script (.dat file extension) used to calculate the interaction energy counterpoise corrected (in kcal $\cdot$ mol^{-1}) of a dimer formed by methane molecules bu supermolecular methods (at MP2/aug-cc-pVTZ theory level) using psi4 software [91].

```
molecule ch4_dimer {
   C 0.00000000 0.00000000 0.00000000
   H 0.00000000 1.08452710 0.00000000
   H -0.51125100 -0.36150900 0.88551270
   H -0.51125100 -0.36150900 -0.88551270
   H 1.02250190 -0.36150900 0.00000000
   --
   C 0.00000000 3.77900000 -0.00000000
   H 0.00000000 2.69447290 -0.00000000
   H -0.51125100 4.14050900 -0.88551270
   H -0.51125100 4.14050900 0.88551270
   H 1.02250190 4.14050900 0.00000000
}

set basis aug-cc-pVTZ

E = energy('sapt2+(3)dmp2')
Efinal = E* psi_hartree2kcalmol
psi4.print_out("sapt2+(3)dmp2/aug-cc-pVTZ = ")
psi4.print_out("%10.6f" % (Efinal))
```

Source code 3 – Example of script (.dat file extension) used to calculate the interaction energy (in kcal $\cdot$ mol^{-1}) of a dimer formed by methane molecules by perturbational methods (at SAPT2+(3)δMP2/aug-cc-pVTZ level) using psi4 software [91].

APPENDIX B

SAFT-VR Mie for mixtures

Description: this Appendix briefly presents the SAFT-VR Mie equation of state for mixtures (*i.e.*, the general case). On the following pages of this Appendix, the development of the equations for ideal gas and monomer contributions to the Helmholtz free energy are demonstrated, as well as the dependence of each contribution in relation to the Mie potential parameters. For a more detailed explanation, it is suggested reading the original publication by Lafitte *et al.* [87].

Ideal contribution

The ideal contribution to the Helmholtz free energy is expressed by Eq. (B.1):

$$a^{\text{ideal}} = \left(\sum_{i=1}^{n} x_i \ln \rho_i \Lambda_i^3 \right) - 1, \tag{B.1}$$

where n is the number of species in the mixture, x_i is the mole fraction of the specie i in the mixture (with $x_i = N_i/N$, where N_i is the number of molecules of specie i in the mixture and N is the total number of molecules), ρ_i is the density of specie i in the mixture (with $\rho_i = N_i/V$, where N_i is the number of molecules of specie i in the mixture, and V is the total volume of the mixture), and Λ_i is the de Broglie thermal wavelength of species i in the mixture.

Monomer contribution

The monomer contribution to the Helmholtz free energy is expressed by Eq. (B.2):

$$a^{\text{mono}} = \left(\sum_{i=1}^{n} x_i m_i \right) a^{\text{M}}, \tag{B.2}$$

where m_i is the number of monomers in the chain i, a^{M} is the monomer Helmholtz free energy.

The expression for monomer Helmholtz free energy (a^{M}) is given by Eq. (B.3):

$$a^{\text{M}} = a^{\text{HS}} + \beta a_1 + (\beta)^2 a_2 + (\beta)^3 a_3, \tag{B.3}$$

where a^{HS} is a expression for a multicomponent mixture of hard spheres, $\beta = 1/(k_B T)$ (with k_B the Boltzmann constant and T the absolute temperature), and a_1, a_2, and a_3 are perturbation terms of first, second, and third-order, respectively.

In Eq. (B.3), the expression for a multicomponent mixture of hard spheres (a^{HS}) is given by Eq. (B.4):

$$a^{\text{HS}} = \frac{6}{\pi \rho_s} \left[\left(\frac{\zeta_2^3}{\zeta_3^2} - \zeta_0 \right) \ln(1-\zeta_3) + \frac{3\zeta_1 \zeta_2}{(1-\zeta_3)} + \frac{\zeta_2^3}{\zeta_3 (1-\zeta_3)^2} \right], \tag{B.4}$$

with the terms obtained through the set of Eqs. (B.5)-(B.7)

$$\zeta_l = \frac{\pi \rho_s}{6} \left(\sum_{i=1}^{n} x_{s,i} d_{ii}^l \right), \quad \text{for } l = 0,1,2,3; \tag{B.5}$$

$$x_{s,i} = \frac{m_i x_i}{\sum_{k=1}^{n} m_k x_k}, \tag{B.6}$$

$$d_{ii} = \int_0^{\sigma_{ii}} \left(1 - \exp\left(-\beta U_{ii}^{\text{Mie}}(r)\right)\right) \mathrm{d}r, \tag{B.7}$$

where ρ_s is the total number density of spherical segments, ζ_l is the moment of order l to the number density, $x_{s,i}$ is the mole fraction of segments of component i, d_{ii} is the effective diameter for the segments of each component, and $U_{ii}^{\text{Mie}}(r)$ is the equation for the Mie potential (see Eq. (2.100) and (2.101) for details).

In Eq. (B.3), the expression for the first-order perturbation term (a_1) is given by Eq. (B.8):

$$a_1 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{s,i}x_{s,j}a_{1,ij}, \tag{B.8}$$

with the terms obtained through the set of Eqs. (B.9)-(B.16)

$$\begin{aligned} a_{1,ij} = \; & C_{ij}\Big[x_{0,ij}^{\lambda_{\text{att},ij}}\left(a_{1,ij}^{S}\left(\rho_s;\lambda_{\text{att},ij}\right) + B_{ij}\left(\rho_s;\lambda_{\text{att},ij}\right)\right) \\ & -x_{0,ij}^{\lambda_{\text{rep},ij}}\left(a_{1,ij}^{S}\left(\rho_s;\lambda_{\text{rep},ij}\right) + B_{ij}\left(\rho_s;\lambda_{\text{rep},ij}\right)\right)\Big], \end{aligned} \tag{B.9}$$

$$B_{ij}\left(\rho_s;\lambda_{ij}\right) = 2\pi\rho_s d_{ij}^3\varepsilon_{ij}\left[\frac{1-\zeta_x/2}{(1-\zeta_x)}I_{\lambda,ij} - \frac{9\zeta_x(1+\zeta_x)}{2(1-\zeta_x)^3}J_{\lambda,ij}\right], \tag{B.10}$$

$$\zeta_x = \frac{\pi\rho_s}{6}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{s,i}x_{s,j}d_{ij}^3, \tag{B.11}$$

$$I_{\lambda,ij} = -\frac{x_{0,ij}^{3-\lambda_{k,ij}} - 1}{\lambda_{k,ij} - 3}, \quad \text{for } k = \text{att, rep;} \tag{B.12}$$

$$J_{\lambda,ij} = -\frac{\left(x_{0,ij}\right)^{4-\lambda_{k,ij}}\left(\lambda_{k,ij}-3\right) - \left(x_{0,ij}\right)^{3-\lambda_{k,ij}}\left(\lambda_{k,ij}-4\right) - 1}{\left(\lambda_{k,ij}-3\right)\left(\lambda_{k,ij}-4\right)}, \quad \text{for } k = \text{att, rep;} \tag{B.13}$$

$$a_{1,ij}^{s}\left(\rho_s;\lambda_{k,ij}\right) = -2\rho_s\left(\frac{\pi\varepsilon_{ij}d_{ij}^3}{\lambda_{k,ij}-3}\right)\frac{1-\zeta_x^{\text{eff}}\left(\lambda_{k,ij}\right)/2}{\left(1-\zeta_x^{\text{eff}}\left(\lambda_{k,ij}\right)\right)^3}, \quad \text{for } k = \text{att, rep;} \tag{B.14}$$

$$\zeta_x^{\text{eff}}\left(\lambda_{k,ij}\right) = c_1\left(\lambda_{k,ij}\right)\zeta_x + c_2\left(\lambda_{k,ij}\right)\zeta_x^2 + c_3\left(\lambda_{k,ij}\right)\zeta_x^3 + c_4\left(\lambda_{k,ij}\right)\zeta_x^4, \quad \text{for } k = \text{att, rep;} \tag{B.15}$$

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = \begin{pmatrix} 0.81096 & 1.7888 & -37.578 & 92.284 \\ 1.0205 & -19.341 & 151.25 & -463.50 \\ -1.9057 & 22.845 & -228.14 & 973.92 \\ 1.0885 & -6.1962 & 106.98 & -677.64 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1/\lambda_{k,ij} \\ 1/\lambda_{k,ij}^2 \\ 1/\lambda_{k,ij}^3 \end{pmatrix}, \quad \text{for } k = \text{att, rep;} \tag{B.16}$$

where C_{ij}, $\lambda_{\text{att},ij}$, and $\lambda_{\text{rep},ij}$ are parameters of the Mie potential (see Eq. (2.100) and (2.101) for details); $x_{0,ij} = \sigma_{ij}/d_{ij}$ (with σ_{ij} referring to the Mie potential parameter, see Eq. (2.100)

and (2.101) for details) , ζ_x^{eff} is the effective packing fraction, ζ_x is the packing fraction of the mixture.

In Eq. (B.3), the expression for the second-order perturbation term (a_2) is given by Eq. (B.17):

$$a_2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{s,i}x_{s,j}a_{2,ij}, \tag{B.17}$$

with the terms obtained through the set of Eqs. (B.18)-(B.23)

$$\begin{aligned} a_{2,ij} =& \frac{1}{2}K_{\text{HS}}\left(1+\chi_{ij}\right)\varepsilon_{ij}C_{ij}^2 \times \left[x_{0,ij}^{2\lambda_{\text{att},ij}}\left(a_{1,ij}^S\left(\rho_s;2\lambda_{\text{att},ij}\right)\right) + B_{ij}\left(\rho_s;2\lambda_{\text{att},ij}\right)\right. \\ & - 2x_{0,ij}^{\lambda_{\text{att},ij}+\lambda_{\text{rep},ij}}\left(a_{1,ij}^S\left(\rho_s;\lambda_{\text{att},ij}+\lambda_{\text{rep},ij}\right) + B_{ij}\left(\rho_s;\lambda_{\text{att},ij}+\lambda_{\text{rep},ij}\right)\right) \\ & \left. + x_{0,ij}^{2\lambda_{\text{rep},ij}}\left(a_{1,ij}^S\left(\rho_s;2\lambda_{\text{rep},ij}\right) + B_{ij}\left(\rho_s;2\lambda_{\text{rep},ij}\right)\right)\right], \end{aligned} \tag{B.18}$$

$$K^{\text{HS}} = \frac{\left(1-\zeta_x\right)^4}{1+4\zeta_x+4\zeta_x^2-4\zeta_x^3+\zeta_x^4}, \tag{B.19}$$

$$\chi_{ij} = f_1\left(\alpha_{ij}\right)\bar{\zeta}_x + f2\left(\alpha_{ij}\right)\bar{\zeta}_x^5 + f_3\left(\alpha_{ij}\right)\bar{\zeta}_x^8, \tag{B.20}$$

$$\bar{\zeta}_x = \frac{\pi\rho_s}{6}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{s,i}x_{s,j}\sigma_{ij}^3, \tag{B.21}$$

$$\alpha_{ij} = C_{ij}\left(\frac{1}{\lambda_{\text{att},ij}-3} - \frac{1}{\lambda_{\text{rep},ij}-3}\right), \tag{B.22}$$

where K^{HS} is the isothermal compressibility of the mixture of hard spheres and a_{ij} dimensionless van der Waals-like integrated attractive energy.

In Eq. (B.3), the expression for the third-order perturbation term (a_3) is given by Eq. (B.23):

$$a_{3,ij} = -\varepsilon_{ij}^3 f_4\left(\alpha_{ij}\right)\bar{\zeta}_x \exp\left(f_5\left(\alpha_{ij}\right)\bar{\zeta}_x + f_6\left(\alpha_{ij}\right)\bar{\zeta}_x^2\right), \tag{B.23}$$

with the terms obtained through the Eq. (B.24)

$$f_k\left(\alpha_{ij}\right) = \left(\sum_{n=0}^{n=3}\phi_{k,n}\alpha_{ij}^n\right) / \left(1+\sum_{n=4}^{n=6}\phi_{k,n}\alpha_{ij}^{n-3}\right), \tag{B.24}$$

where α_{ij} coefficients are also given by Eq. (B.23).

APPENDIX C

NIST reference data

Description: this Appendix describes the constitution of the datasets of experimental data used in this work and which were taken from NIST [95]. On the following pages, the temperature and pressure conditions and the number of data points considered to build the datasets to be used in AARD% calculations of thermophysical and saturation properties are presented in tables.

Table C1 – Isobars, temperature ranges and number of data points obtained from NIST to calculate AARD% values for thermophysical properties of 26 molecules studied in this work.

Molecule	Isobars	Temperature range	Number of data points
helium	10, 50, and 100 MPa	5 to 750 K	447
neon	10, 50, and 100 MPa	30 to 725 K	417
argon	10, 50, and 100 MPa	90 to 750 K	393
krypton	10, 50, and 100 MPa	120 to 750 K	373
methane	50, 100, 250, 500, and 1000 atm	120 to 600 K	105
tetrafluoromethane	50, 100, 250, and 500 atm	155 to 600 K	48
fluoromethane	50, 100, 250, and 500 atm	220 to 425 K	42
chlorotrifluoromethane	50, 100, 250, and 300 atm	180 to 400 K	48
dichlorodifluoromethane	50, 100, 250, and 500 atm	265 to 445 K	36
dichlorofluoromethane	50, 100, 250, and 500 atm	270 to 472 K	42
chlorodifluoromethane	50, 100, 250, and 500 atm	240 to 550 K	48
hydrogen	10, 50, and 100 MPa	20 to 800 K	466
nitrogen	10, 50, and 100 MPa	70 to 800 K	437
oxygen	10, 20, and 40 MPa	60 to 800 K	447
fluorine	10, 15, and 20 MPa	55 to 300 K	296
carbon monoxide	10, 50, and 100 MPa	75 to 500 K	254
carbon dioxide	10, 50, and 100 MPa	220 to 800 K	345
ethane	50, 100, and 200 MPa	100 to 550 K	264
propane	50, 100, and 200 MPa	95 to 550 K	274
butane	50, 100, and 200 MPa	145 to 550 K	240
isobutane	50, 100, and 200 MPa	130 to 550 K	239
ethene	50, 100, and 200 MPa	115 to 450 K	200
propene	50, 100, and 200 MPa	210 to 550 K	172
propyne	10, 20, and 30 MPa	346 to 450 K	176
cyclopropane	10, 15, and 20 MPa	286 to 450 K	242
benzene	50, 100, and 200 MPa	315 to 700 K	201

Table C2 – Temperature ranges and number of data points obtained from NIST to calculate AARD% values for thermophysical properties of 26 molecules studied in this work.

Molecule	Temperature range	Number of data points
helium	2.2 to 5 K	15
neon	25 to 43 K	48
argon	84 to 145 K	62
krypton	116 to 208 K	47
methane	92 to 172 K	41
tetrafluoromethane	120 to 222 K	52
fluoromethane	130 to 307 K	60
chlorotrifluoromethane	100 to 295 K	66
dichlorodifluoromethane	150 to 378 K	77
dichlorofluoromethane	150 to 366 K	73
chlorodifluoromethane	200 to 360 K	41
hydrogen	14 to 32 K	37
nitrogen	64 to 123 K	60
oxygen	56 to 150 K	64
fluorine	54 to 113 K	60
carbon monoxide	70 to 130 K	61
carbon dioxide	218 to 297 K	54
ethane	101 to 299 K	67
propane	122 to 362 K	61
butane	152 to 416 K	67
isobutane	142 to 398 K	65
ethene	110 to 275 K	56
propene	130 to 358 K	58
propyne	276 to 400 K	63
cyclopropane	276 to 390 K	58
benzene	300 to 548 K	63

APPENDIX D

Results presented in tables

Description: this Appendix presents in tables most of the numerical results previously shown in figures throughout this dissertation. On the following pages of this Appendix, tables referring to AARD% values for thermophysical properties, percentage deviations for critical point, and AARD% values for saturation properties in relation to NIST reference data [95] are presented.

Table D1 – AARD% values for thermophysical properties for some noble gases using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state.

			AARD% values					
Molecule	Methodology[a]	Context[b]	ρ	c_v	c_p	c_{sound}	μ_{JT}	Reference
helium	TD	SAFT framework	14.65	4.73	5.66	15.40	89.97	[104][c]
	BU (S)	DLPNO-CCSD(T)/5Z	1.65	4.45	6.67	1.74	9.35	[10][c]
		MP2/aug-cc-pVDZ (A)	1.80	4.38	5.05	2.08	11.14	this work
		MP2/aug-cc-pVTZ (A)	1.51	4.20	5.11	1.79	9.65	this work
		MP2/aug-cc-pVQZ (A)	1.71	4.07	5.20	2.00	10.66	this work
		MP4/aug-cc-pVDZ (A)	3.47	3.73	5.03	3.41	20.63	this work
		MP4/aug-cc-pVTZ (A)	3.12	3.94	6.08	3.83	14.90	this work
		MP4/aug-cc-pVQZ (A)	2.77	4.08	6.28	3.53	12.75	this work
		CCSD(T)/aug-cc-pVDZ (A)	3.44	3.75	5.10	3.43	20.19	this work
		CCSD(T)/aug-cc-pVTZ (A)	3.44	4.13	6.75	4.47	15.90	this work
		CCSD(T)/aug-cc-pVQZ (A)	2.76	4.14	6.43	3.58	12.45	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	109.83	2.41	2.96	23.91	28.56	this work
		SAPT2+/aug-cc-pVDZ (A)	3.90	4.46	5.88	4.63	21.18	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	2.10	4.58	6.43	2.95	10.98	this work
neon	TD	SAFT framework	1.11	1.08	0.73	1.34	8.57	[103][c]
	BU (S)	DLPNO-CCSD(T)/5Z	0.70	1.41	1.52	0.96	19.48	[10][c]
		MP2/aug-cc-pVDZ (A)	9.63	1.61	6.59	4.62	165.94	this work
		MP2/aug-cc-pVTZ (A)	6.70	1.20	4.84	3.07	109.30	this work
		MP2/aug-cc-pVQZ (A)	5.35	1.21	3.68	2.56	83.37	this work
		MP4/aug-cc-pVDZ (A)	8.96	1.52	5.99	4.28	149.26	this work
		MP4/aug-cc-pVTZ (A)	3.94	1.09	3.02	1.85	61.25	this work
		MP4/aug-cc-pVQZ (A)	2.65	1.00	1.61	1.56	33.71	this work
		CCSD(T)/aug-cc-pVDZ (A)	10.16	1.44	6.30	4.99	168.39	this work
		CCSD(T)/aug-cc-pVTZ (A)	3.60	1.09	2.91	1.73	57.52	this work
		CCSD(T)/aug-cc-pVQZ (A)	2.73	0.99	1.58	1.68	33.83	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	11.70	1.59	7.20	5.82	204.19	this work
		SAPT2+/aug-cc-pVDZ (A)	8.76	1.55	5.75	4.20	144.17	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	4.06	1.09	3.02	1.92	62.38	this work
argon	TD	SAFT framework	0.72	1.87	1.19	1.21	14.04	[103][c]
		SAFT framework	1.94	1.68	2.38	3.53	45.12	[104][c]
	BU (S)	DLPNO-CCSD(T)/5Z	1.06	1.83	1.41	2.45	21.23	[10][c]
		MP2/aug-cc-pVDZ (A)	13.84	4.15	10.25	7.08	528.66	this work
		MP2/aug-cc-pVTZ (A)	2.92	1.87	2.37	2.29	67.52	this work
		MP2/aug-cc-pVQZ (A)	4.84	2.39	6.55	5.22	142.84	this work
		MP4/aug-cc-pVDZ (A)	17.38	4.99	13.10	8.48	723.87	this work
		MP4/aug-cc-pVTZ (A)	7.68	2.92	5.65	4.12	243.53	this work
		MP4/aug-cc-pVQZ (A)	1.24	1.94	1.89	1.99	30.16	this work
		CCSD(T)/aug-cc-pVDZ (A)	17.53	5.01	13.22	8.53	733.03	this work
		CCSD(T)/aug-cc-pVTZ (A)	8.47	3.15	6.21	4.55	277.17	this work
		CCSD(T)/aug-cc-pVQZ (A)	2.69	1.95	2.74	1.58	71.98	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	29.56	5.43	21.63	13.50	1648.80	this work
		SAPT2+/aug-cc-pVDZ (A)	16.48	4.88	12.60	8.15	677.71	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	8.00	2.99	5.84	4.29	256.26	this work
krypton	TD	SAFT framework	0.84	3.31	1.52	1.56	36.38	[103][c]
	BU (S)	DLPNO-CCSD(T)/5Z	1.65	3.18	1.72	2.45	54.05	[10][c]
		MP2/aug-cc-pVDZ (A)	19.00	6.48	14.93	10.08	1014.62	this work
		MP2/aug-cc-pVTZ (A)	1.83	3.02	2.21	3.73	45.70	this work
		MP2/aug-cc-pVQZ (A)	13.21	4.74	16.26	13.03	415.17	this work
		MP4/aug-cc-pVDZ (A)	23.91	7.67	18.68	11.93	1396.79	this work
		MP4/aug-cc-pVTZ (A)	10.37	3.82	7.79	5.41	448.68	this work
		MP4/aug-cc-pVQZ (A)	1.43	2.87	3.95	4.46	31.50	this work
		CCSD(T)/aug-cc-pVDZ (A)	24.21	7.70	18.98	12.01	1424.60	this work
		CCSD(T)/aug-cc-pVTZ (A)	12.36	4.55	9.39	6.62	566.26	this work
		CCSD(T)/aug-cc-pVQZ (A)	3.70	2.88	3.85	2.01	151.53	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	31.44	7.96	26.99	14.24	2278.16	this work
		SAPT2+/aug-cc-pVDZ (A)	23.21	7.60	18.38	11.71	1345.30	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	10.76	3.95	8.07	5.64	470.43	this work

[a]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods, and BU (P) refers to bottom-up methodology and perturbational methods. [b](A) indicates that approach A objective function was minimized in the parameter fitting procedure. [c]AARD% values calculated for the datasets of this work.

Table D2 – Critical point and percentage deviation from NIST reference data computed using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state for some noble gases.

			Value			Deviation from NIST (%)			
Molecule	Methodology[a]	Context[b]	T_{crit} / K	p_{crit} / atm	ρ_{crit} / kg·m^{-3}	T_{crit}	p_{crit}	ρ_{crit}	Reference
helium	experimental	-	5.20	0.23	69.58	-	-	-	[95]
	TD	SAFT framework	5.19	0.19	54.28	0.09	16.17	21.99	[104][c]
	BU (S)	DPLNO-CCSD(T)/5Z	10.68	0.75	101.18	105.55	229.69	45.42	[10][c]
		MP2/aug-cc-pVDZ (A)	10.72	0.61	80.72	106.30	167.10	16.01	this work
		MP2/aug-cc-pVTZ (A)	10.67	0.63	84.16	105.45	176.87	20.95	this work
		MP2/aug-cc-pVQZ (A)	10.48	0.63	86.06	101.76	177.18	23.68	this work
		MP4/aug-cc-pVDZ (A)	7.14	0.43	87.05	37.52	87.16	25.10	this work
		MP4/aug-cc-pVTZ (A)	9.50	0.63	96.71	82.85	174.60	39.00	this work
		MP4/aug-cc-pVQZ (A)	10.34	0.69	98.20	99.12	204.31	41.14	this work
		CCSD(T)/aug-cc-pVDZ (A)	7.34	0.44	87.54	41.19	93.23	25.81	this work
		CCSD(T)/aug-cc-pVTZ (A)	10.36	0.71	101.37	99.36	212.42	45.69	this work
		CCSD(T)/aug-cc-pVQZ (A)	10.72	0.72	98.80	106.31	217.18	41.99	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	1.12	0.04	54.56	78.39	81.51	21.59	this work
		SAPT2+/aug-cc-pVDZ (A)	10.82	0.62	82.83	108.18	172.39	19.04	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	13.20	0.84	91.30	154.13	268.65	31.21	this work
neon	experimental	-	44.40	2.66	486.00	-	-	-	[95]
	TD	SAFT framework	45.25	2.81	446.62	1.92	5.68	8.10	[103][c]
	BU (S)	DPLNO-CCSD(T)/5Z	50.56	3.07	439.26	13.88	15.32	9.62	[10][c]
		MP2/aug-cc-pVDZ (A)	13.51	0.64	347.60	69.56	76.09	28.48	this work
		MP2/aug-cc-pVTZ (A)	23.18	1.19	374.62	47.80	55.23	22.92	this work
		MP2/aug-cc-pVQZ (A)	28.49	1.58	407.11	35.83	40.59	16.23	this work
		MP4/aug-cc-pVDZ (A)	16.68	0.80	354.72	62.44	69.79	27.01	this work
		MP4/aug-cc-pVTZ (A)	32.57	1.86	417.81	26.65	29.99	14.03	this work
		MP4/aug-cc-pVQZ (A)	39.46	2.40	447.83	11.14	9.98	7.85	this work
		CCSD(T)/aug-cc-pVDZ (A)	14.85	0.67	334.23	66.55	74.66	31.23	this work
		CCSD(T)/aug-cc-pVTZ (A)	33.09	1.94	428.95	25.48	27.17	11.74	this work
		CCSD(T)/aug-cc-pVQZ (A)	39.64	2.42	451.77	10.71	8.97	7.04	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	9.44	0.39	306.75	78.74	85.26	36.88	this work
		SAPT2+/aug-cc-pVDZ (A)	17.77	0.87	359.60	59.97	67.38	26.01	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	32.48	1.85	415.83	26.85	30.51	14.44	this work
argon	experimental	-	150.69	4.86	535.60	-	-	-	[95]
	TD	SAFT framework	153.88	5.36	504.94	2.12	10.20	5.73	[103][c]
		SAFT framework	154.31	5.39	513.25	2.40	10.86	4.17	[104][c]
	BU (S)	DPLNO-CCSD(T)/5Z	159.50	5.42	492.02	5.85	11.55	8.14	[10][c]
		MP2/aug-cc-pVDZ (A)	97.55	2.96	442.84	35.26	39.20	17.32	this work
		MP2/aug-cc-pVTZ (A)	149.01	5.11	501.77	1.11	5.00	6.32	this work
		MP2/aug-cc-pVQZ (A)	176.46	6.40	529.36	17.11	31.51	1.16	this work
		MP4/aug-cc-pVDZ (A)	77.14	2.22	421.90	48.81	54.26	21.23	this work
		MP4/aug-cc-pVTZ (A)	127.48	4.22	484.75	15.40	13.28	9.49	this work
		MP4/aug-cc-pVQZ (A)	153.25	5.38	513.26	1.70	10.63	4.17	this work
		CCSD(T)/aug-cc-pVDZ (A)	76.21	2.19	420.72	49.43	54.94	21.45	this work
		CCSD(T)/aug-cc-pVTZ (A)	123.66	4.06	481.44	17.94	16.46	10.11	this work
		CCSD(T)/aug-cc-pVQZ (A)	146.97	5.11	508.22	2.47	5.05	5.11	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	14.75	0.28	282.95	90.21	94.16	47.17	this work
		SAPT2+/aug-cc-pVDZ (A)	81.07	2.39	430.50	46.20	50.94	19.62	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	126.12	4.15	482.75	16.30	14.56	9.87	this work
krypton	experimental	-	209.48	5.53	909.20	-	-	-	[95]
	TD	SAFT framework	214.26	6.12	868.01	2.28	10.77	4.53	[103][c]
	BU (S)	DPLNO-CCSD(T)/5Z	214.46	6.03	858.77	2.38	9.19	5.55	[10][c]
		MP2/aug-cc-pVDZ (A)	126.90	3.09	747.08	39.42	43.99	17.83	this work
		MP2/aug-cc-pVTZ (A)	217.75	6.10	858.29	3.95	10.37	5.60	this work
		MP2/aug-cc-pVQZ (A)	268.17	8.06	921.33	28.02	45.90	1.33	this work
		MP4/aug-cc-pVDZ (A)	94.97	2.17	700.92	54.66	60.76	22.91	this work
		MP4/aug-cc-pVTZ (A)	176.97	4.72	819.55	15.52	14.63	9.86	this work
		MP4/aug-cc-pVQZ (A)	223.61	6.45	887.25	6.74	16.80	2.41	this work
		CCSD(T)/aug-cc-pVDZ (A)	92.66	2.10	697.13	55.77	61.92	23.32	this work
		CCSD(T)/aug-cc-pVTZ (A)	166.24	4.37	808.34	20.64	20.91	11.09	this work
		CCSD(T)/aug-cc-pVQZ (A)	205.28	5.81	870.68	2.00	5.22	4.24	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	32.23	0.58	548.68	84.61	89.54	39.65	this work
		SAPT2+/aug-cc-pVDZ (A)	98.02	2.27	711.64	53.21	58.86	21.73	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	175.02	4.65	817.44	16.45	15.80	10.09	this work

[a]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods, and BU (P) refers to bottom-up methodology and perturbational methods. [b](A) indicates that approach A objective function was minimized in the parameter fitting procedure. [c]Percentage deviation values recalculated in this work.

Table D3 – AARD% values for saturation properties for some noble gases using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state.

			AARD% values		
Molecule	Methodology[a]	Context[b]	p_{sat}	ΔH_{vap}	Reference
argon	TD	SAFT framework	0.35	4.30	[103][c]
		SAFT framework	7.92	10.20	[104][c]
	BU (S)	DLPNO-CCSD(T)/5Z	21.21	12.90	[10][c]
		MP2/aug-cc-pVTZ (A)	15.31	1.75	this work
krypton	TD	SAFT framework	0.47	8.90	[103][c]
	BU (S)	DLPNO-CCSD(T)/5Z	3.89	10.80	[10][c]
		MP2/aug-cc-pVTZ (A)	14.37	18.09	this work

[a]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods. [b](A) indicates that approach A objective function was minimized in the parameter fitting procedure. [c]AARD% values calculated for the datasets of this work.

Table D4 – AARD% values for thermophysical properties for methane using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state and molecular dynamics simulations.

			AARD% values				
Calculation method[a]	Methodology[b]	Context[c]	ρ	c_v	c_p	c_{sound}	Reference
SAFT-VR Mie EoS	TD	SAFT framework	0.49	1.40	1.30	1.25	[103][e]
	BU (S)	DLPNO-CCSD(T)/5Z	6.28	1.38	3.12	2.93	[10][e]
		MP2/aug-cc-pVDZ (A)	11.58	2.88	9.69	8.61	this work
		MP2/aug-cc-pVDZ (B)	10.73	2.16	5.54	7.88	this work
		MP2/aug-cc-pVTZ (A)	1.46	1.45	2.62	1.00	this work
		MP2/aug-cc-pVTZ (B)	27.93	2.29	5.98	12.96	this work
		MP2/aug-cc-pVQZ (A)	1.43	1.46	1.16	1.57	this work
		MP2/aug-cc-pVQZ (B)	34.01	2.80	13.54	16.37	this work
		MP4/aug-cc-pVDZ (A)	5.78	1.76	8.25	4.13	this work
		MP4/aug-cc-pVDZ (B)	17.13	1.92	4.94	7.25	this work
		CCSD(T)/aug-cc-pVDZ (A)	6.69	1.90	5.82	4.81	this work
		CCSD(T)/aug-cc-pVDZ (B)	15.87	1.82	4.51	6.76	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	25.53	5.63	16.07	16.40	this work
		sSAPT0/jun-cc-pVDZ (B)	19.50	6.36	16.75	17.90	this work
		SAPT2+/aug-cc-pVDZ (A)	0.88	1.45	1.04	1.46	this work
		SAPT2+/aug-cc-pVDZ (B)	39.43	2.83	11.67	17.59	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	5.88	1.58	2.28	5.07	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (B)	36.83	3.25	9.31	19.29	this work
MD simulations	TD	SAFT framework	2.37	1.58	2.85	3.60	[d,e]
	BU (S)	MP2/aug-cc-pVDZ (A)	11.27	3.01	9.33	8.54	this work
		MP2/aug-cc-pVDZ (B)	10.92	2.23	4.66	8.13	this work
		MP2/aug-cc-pVTZ (A)	1.07	1.42	1.70	1.15	this work
		MP2/aug-cc-pVTZ (B)	28.59	0.97	5.98	12.49	this work
		MP2/aug-cc-pVQZ (A)	1.85	0.95	1.43	1.58	this work
		MP2/aug-cc-pVQZ (B)	35.13	1.43	9.95	16.06	this work
		MP4/aug-cc-pVDZ (A)	5.32	2.18	7.63	4.15	this work
		MP4/aug-cc-pVDZ (B)	17.67	1.07	4.13	7.06	this work
		CCSD(T)/aug-cc-pVDZ (A)	6.34	2.15	6.33	4.82	this work
		CCSD(T)/aug-cc-pVDZ (B)	16.37	1.27	4.50	6.53	this work
	BU (P)	sSAPT0/jun-cc-pVDZ (A)	25.39	5.54	16.01	16.42	this work
		sSAPT0/jun-cc-pVDZ (B)	19.36	6.32	16.68	17.99	this work
		SAPT2+/aug-cc-pVDZ (A)	1.32	0.96	1.63	1.57	this work
		SAPT2+/aug-cc-pVDZ (B)	34.68	1.40	9.05	16.19	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (A)	6.42	0.66	2.69	5.05	this work
		SAPT2+(3)δMP2/aug-cc-pVTZ (B)	37.73	1.72	9.63	18.89	this work

[a]SAFT-VR Mie EoS refers to SAFT-VR Mie equation of state and MD simulations refers to molecular dynamics simulations. [b]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods, and BU (P) refers to bottom-up methodology and perturbational methods. [c](A) and (B) indicate that approaches A and B objective functions were minimized in the parameter fitting procedure, respectively. [d]Parameters obtained through personal communication with professor Erich A. Müller (June 3rd, 2021), based on published works [106, 107]. [e]AARD% values calculated for the datasets of this work.

Table D5 – Critical point and percentage deviation from NIST reference data computed using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state for methane.

		Value			Deviation from NIST (%)			
Methodology[a]	Context[b]	T_{crit} / K	p_{crit} / atm	ρ_{crit} / kg·m^{-3}	T_{crit}	p_{crit}	ρ_{crit}	Reference
experimental	-	190.56	45.39	162.66	-	-	-	[95]
TD	SAFT framework	190.53	49.79	156.14	0.02	9.69	4.01	[c,d]
	SAFT framework	195.54	50.85	153.70	2.61	12.02	5.51	[103][d]
BU (S)	DLPNO-CCSD(T)/5Z	183.02	43.52	140.36	3.96	4.11	13.71	[10][d]
	MP2/aug-cc-pVDZ (A)	152.24	37.27	144.49	20.11	17.88	11.17	this work
	MP2/aug-cc-pVDZ (B)	186.56	63.30	200.03	2.10	39.45	22.97	this work
	MP2/aug-cc-pVTZ (A)	189.10	49.67	155.09	0.77	9.42	4.65	this work
	MP2/aug-cc-pVTZ (B)	196.54	51.62	155.09	3.14	13.73	4.65	this work
	MP2/aug-cc-pVQZ (A)	198.98	53.09	157.56	4.41	16.96	3.13	this work
	MP2/aug-cc-pVQZ (B)	247.16	91.96	219.36	29.70	102.60	34.86	this work
	MP4/aug-cc-pVDZ (A)	175.84	44.55	149.56	7.73	1.85	8.06	this work
	MP4/aug-cc-pVDZ (B)	215.77	75.08	205.15	13.23	65.42	26.12	this work
	CCSD(T)/aug-cc-pVDZ (A)	172.53	43.48	148.74	9.46	4.22	8.56	this work
	CCSD(T)/aug-cc-pVDZ (B)	211.68	73.37	204.36	11.08	61.65	25.63	this work
BU (P)	sSAPT0/jun-cc-pVDZ (A)	76.42	16.35	126.26	59.90	63.98	22.38	this work
	sSAPT0/jun-cc-pVDZ (B)	75.04	21.21	166.61	60.62	53.28	2.43	this work
	SAPT2+/aug-cc-pVDZ (A)	198.45	52.34	155.69	4.14	15.31	4.28	this work
	SAPT2+/aug-cc-pVDZ (B)	248.02	91.12	216.60	30.15	100.75	33.16	this work
	SAPT2+(3)δMP2/aug-cc-pVTZ (A)	209.41	56.23	158.59	9.89	23.88	2.50	this work
	SAPT2+(3)δMP2/aug-cc-pVTZ (B)	256.79	94.46	216.88	34.75	108.11	33.33	this work

[a]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods, and BU (P) refers to bottom-up methodology and perturbational methods. [b](A) and (B) indicate that approaches A and B objective functions were minimized in the parameter fitting procedure, respectively. [c]Parameters obtained through personal communication with professor Erich A. Müller (June 3rd, 2021), based on published works [106, 107]. [d]Percentage deviation values recalculated in this work.

Table D6 – AARD% values for saturation properties for methane using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state.

		AARD% values		
Methodology[a]	Context[b]	p_{sat}	ΔH_{vap}	Reference
TD	SAFT framework	0.46	2.84	[103][c]
	SAFT framework	7.75	3.76	[104][c]
BU (S)	DLPNO-CCSD(T)/5Z	71.99	16.40	[10][c]
	MP2/aug-cc-pVTZ (A)	28.39	4.23	this work

[a]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods. [b](A) indicates that approach A objective function was minimized in the parameter fitting procedure. [c]AARD% values calculated for the datasets of this work.

Table D7 – AARD% values for thermophysical properties for some substituted-methane compounds of this study using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state and molecular dynamics simulations.

				AARD% values				
Calculation method[a]	Molecule	Methodology[b]	Context[c]	ρ	c_v	c_p	c_{sound}	Reference
SAFT-VR Mie EoS	tetrafluoromethane	BU (S)	MP2/aug-cc-pVTZ (A)	3.26	2.93	5.16	2.46	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	92.49	5.25	15.37	41.63	this work
	fluoromethane	BU (S)	MP2/aug-cc-pVTZ (A)	21.46	108.29	109.27	12.41	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	341.39	258.51	120.39	437.32	this work
	chlorotrifluoromethane	BU (S)	MP2/aug-cc-pVTZ (A)	17.88	188.50	214.85	22.96	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	290.81	224.57	155.86	231.94	this work
	dichlorodifluoromethane	BU (S)	MP2/aug-cc-pVTZ (A)	41.36	101.64	89.07	66.42	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	261.79	144.45	79.00	293.31	this work
	chlorodifluoromethane	BU (S)	MP2/aug-cc-pVTZ (A)	16.25	161.16	181.33	20.17	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	255.79	198.79	132.51	198.85	this work
	dichlorofluoromethane	BU (S)	MP2/aug-cc-pVTZ (A)	5.30	99.54	80.87	26.57	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	175.57	183.28	84.67	253.83	this work
MD simulations	tetrafluoromethane	TD	SAFT framework	2.83	4.05	1.95	4.99	[d,e]
		BU (S)	MP2/aug-cc-pVTZ (A)	3.58	3.07	6.77	2.48	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	94.02	2.78	15.52	40.40	this work
	fluoromethane	TD	SAFT framework	9.24	112.64	133.02	22.65	[d,e]
		BU (S)	MP2/aug-cc-pVTZ (A)	20.71	109.75	108.89	12.36	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	362.49	141.80	336.18	478.01	this work
	chlorotrifluoromethane	TD	SAFT framework	4.12	193.41	199.78	5.68	[d,e]
		BU (S)	MP2/aug-cc-pVTZ (A)	20.03	192.77	214.79	25.14	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	294.99	211.64	281.57	227.73	this work
	dichlorodifluoromethane	TD	SAFT framework	4.36	103.40	99.73	5.20	[d,e]
		BU (S)	MP2/aug-cc-pVTZ (A)	42.72	101.94	88.72	70.29	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	284.41	116.66	162.57	321.21	this work
	chlorodifluoromethane	TD	SAFT framework	5.35	164.44	181.12	8.99	[d,e]
		BU (S)	MP2/aug-cc-pVTZ (A)	17.33	163.70	184.94	22.36	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	260.93	179.64	243.38	197.00	this work
	dichlorofluoromethane	TD	SAFT framework	5.19	102.31	75.03	20.60	[d,e]
		BU (S)	MP2/aug-cc-pVTZ (A)	5.42	101.30	81.23	28.65	this work
		BU (S)	MP2/aug-cc-pVTZ (B)	176.52	118.05	207.12	258.12	this work

[a]SAFT-VR Mie EoS refers to SAFT-VR Mie equation of state and MD simulations refers to molecular dynamics simulations. [b]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods. [c](A) and (B) indicate that approaches A and B objective functions were minimized in the parameter fitting procedure, respectively. [d]Parameters obtained through personal communication with professor Erich A. Müller (June 3rd, 2021), based on published works [106, 107]. [e]AARD% values calculated for the datasets of this work.

Table D8 – Critical point and percentage deviation from NIST reference data calculated using top-down and approach A bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state for some substituted-methane compounds of this study.

			Value			Deviation from NIST (%)			
Molecule	Methodology[a]	Context[b]	T_{crit} / K	p_{crit} / atm	ρ_{crit} / kg·m^{-3}	T_{crit}	p_{crit}	ρ_{crit}	Reference
tetrafluoromethane	experimental	-	227.51	37.01	625.66	-	-	-	[95]
	TD	SAFT framework	227.51	39.69	643.40	0.01	7.25	2.83	[c,d]
	BU (S)	MP2/aug-cc-pVTZ (A)	236.07	40.51	562.26	3.76	9.47	10.13	this work
fluoromethane	experimental	-	317.28	58.20	320.00	-	-	-	[95]
	TD	SAFT framework	317.45	76.68	351.29	0.05	31.76	9.78	[c,d]
	BU (S)	MP2/aug-cc-pVTZ (A)	314.23	59.93	241.35	0.96	2.97	24.58	this work
chlorotrifluoromethane	experimental	-	302.00	38.28	583.00	-	-	-	[95]
	TD	SAFT framework	302.04	42.10	604.87	0.01	9.98	3.75	[c,d]
	BU (S)	MP2/aug-cc-pVTZ (A)	338.56	45.21	519.33	12.11	18.10	10.92	this work
dichlorodifluoromethane	experimental	-	385.12	40.82	565.00	-	-	-	[95]
	TD	SAFT framework	384.98	45.07	593.42	0.04	10.42	5.03	[c,d]
	BU (S)	MP2/aug-cc-pVTZ (A)	492.90	57.85	528.55	27.99	41.73	6.45	this work
chlorodifluoromethane	experimental	-	451.48	51.13	526.01	-	-	-	[95]
	TD	SAFT framework	451.68	58.16	569.93	0.04	13.74	8.35	[c,d]
	BU (S)	MP2/aug-cc-pVTZ (A)	533.66	64.24	461.22	18.20	25.63	12.32	this work
dichlorofluoromethane	experimental	-	369.29	49.25	523.84	-	-	-	[95]
	TD	SAFT framework	369.49	57.00	580.55	0.05	15.73	10.83	[c,d]
	BU (S)	MP2/aug-cc-pVTZ (A)	380.18	53.23	450.88	2.95	8.09	13.93	this work

[a]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods. [b](A) indicates that approach A was minimized in the parameter fitting procedure. [c] Parameters obtained through personal communication with professor Erich A. Müller (June 3rd, 2021), based on published works [106, 107]. [d]Percentage deviation values recalculated in this work.

Table D9 – AARD% values for saturation properties for some substituted-methane compounds of this study using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state and molecular dynamics simulations.

			AARD% values		
Molecule	Methodology[b]	Context[c]	p_{sat}	ΔH_{vap}	Reference
tetrafluoromethane	TD	SAFT framework	3.26	2.93	this work [d,e]
	BU (S)	MP2/aug-cc-pVTZ (A)	92.49	5.25	this work
fluoromethane	TD	SAFT framework	21.46	108.29	this work [d,e]
	BU (S)	MP2/aug-cc-pVTZ (A)	341.39	258.51	this work
chlorotrifluoromethane	TD	SAFT framework	17.88	188.50	this work [d,e]
	BU (S)	MP2/aug-cc-pVTZ (A)	290.81	224.57	this work
dichlorodifluoromethane	TD	SAFT framework	41.36	101.64	this work [d,e]
	BU (S)	MP2/aug-cc-pVTZ (A)	261.79	144.45	this work
chlorodifluoromethane	TD	SAFT framework	16.25	161.16	this work [d,e]
	BU (S)	MP2/aug-cc-pVTZ (A)	255.79	198.79	this work
dichlorofluoromethane	TD	SAFT framework	5.30	99.54	this work [d,e]
	BU (S)	MP2/aug-cc-pVTZ (A)	175.57	183.28	this work

[a]SAFT-VR Mie EoS refers to SAFT-VR Mie equation of state and MD simulations refers to molecular dynamics simulations. [b]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods. [c](A) indicates that approach A objective function was minimized in the parameter fitting procedure. [d]Parameters obtained through personal communication with professor Erich A. Müller (June 3rd, 2021), based on published works [106, 107]. [e]AARD% values calculated for the datasets of this work.

Table D10 – AARD% values for thermophysical properties for fifteen molecules of different chemical classes using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state.

			AARD (%)					
Molecule	Methodology[a]	Context[b]	ρ	c_v	c_p	c_{sound}	μ_{JT}	Reference
hydrogen	TD	SAFT framework	0.75	7.43	6.35	1.38	23.64	[99][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.21	7.89	6.97	1.41	17.83	this work
nitrogen	TD	SAFT framework	0.39	0.92	0.70	0.56	10.14	[103][c]
		SAFT framework	3.36	2.04	2.06	4.30	75.48	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	18.05	3.44	11.59	12.89	462.50	this work
oxygen	TD	SAFT framework	0.46	1.66	0.69	1.13	19.41	[103][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	2.60	2.17	4.75	6.04	90.38	this work
fluorine	TD	SAFT framework	0.58	4.84	2.62	3.53	14.72	[103][c]
		SAFT framework	0.65	4.29	2.80	2.83	9.51	[87][c]
		SAFT framework	3.52	8.53	11.35	6.88	29.04	[107][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	18.33	8.30	14.04	15.43	167.89	this work
carbon monoxide	TD	SAFT framework	0.82	1.51	1.27	1.30	6.62	[103][c]
		SAFT framework	3.86	3.45	3.60	5.77	14.83	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	5.75	1.62	2.50	2.88	29.57	this work
carbon dioxide	TD	SAFT framework	2.81	5.54	1.71	5.95	16.08	[103][c]
		SAFT framework	3.58	4.94	2.61	4.47	22.62	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	8.57	5.81	9.27	8.86	70.16	this work
ethane	TD	SAFT framework	0.95	2.10	1.38	3.39	19.11	[103][c]
		SAFT framework	0.45	1.76	1.36	0.93	16.82	[87][c]
		SAFT framework	3.41	8.23	12.59	10.94	27.09	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	2.35	3.70	2.00	4.16	9.24	this work
propane	TD	SAFT framework	0.87	2.39	2.00	3.82	8.95	[103][c]
		SAFT framework	0.53	2.53	2.00	2.50	7.49	[87][c]
		SAFT framework	3.13	13.22	18.98	14.94	18.46	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.44	7.86	5.35	6.18	7.66	this work
n-butane	TD	SAFT framework	0.26	1.07	1.13	0.57	4.24	[103][c]
		SAFT framework	0.34	0.64	1.61	2.05	5.15	[87][c]
		SAFT framework	0.71	0.78	1.76	3.38	7.06	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.63	4.33	3.09	8.37	6.88	this work
isobutane	TD	SAFT framework	0.48	0.82	1.27	1.96	4.85	[103][c]
		SAFT framework	0.91	1.04	1.31	4.48	6.87	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	3.14	4.14	3.20	7.83	6.27	this work
ethene	TD	SAFT framework	0.92	2.63	2.71	1.74	156.23	[103][c]
		SAFT framework	3.37	7.99	10.62	13.51	172.80	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	1.85	8.06	11.06	12.02	253.40	this work
propene	TD	SAFT framework	0.98	1.58	1.13	3.01	24.85	[103][c]
		SAFT framework	4.03	3.51	5.96	8.58	21.66	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	21.41	2.78	1.44	10.98	23.14	this work
propyne	TD	SAFT framework	0.76	10.53	4.01	5.22	104.07	[103][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	22.26	10.39	18.86	64.81	503.32	this work
cyclopropane	TD	SAFT framework	0.96	6.32	3.16	6.03	93.64	[103][c]
		SAFT framework	4.29	7.38	10.71	11.04	197.13	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	13.52	6.59	11.63	51.52	536.41	this work
benzene	TD	SAFT framework	0.84	0.75	2.00	4.02	14.13	[103][c]
		SAFT framework	0.71	1.41	2.24	0.64	11.85	[87][c]
		SAFT framework	0.55	1.29	2.23	1.98	14.15	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	28.26	6.47	3.95	62.88	21.50	this work

[a]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods, and BU (P) refers to bottom-up methodology and perturbational methods. [b](A) indicates that approach A objective function was minimized in the parameter fitting procedure. [c]AARD% values calculated for the datasets of this work.

Table D11 – Critical point and percentage deviation from NIST reference data calculated using top-down and approach A bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state for fifteen molecules of different chemical classes.

			Value			Deviation from NIST (%)			
Molecule	Methodology[a]	Context[b]	T_{crit} / K	p_{crit} / atm	ρ_{crit} / kg·m^{-3}	T_{crit}	p_{crit}	ρ_{crit}	Reference
hydrogen	experimental	-	33.15	1.30	31.26	-	-	-	[95]
	TD	SAFT framework	37.46	1.57	29.76	13.03	20.76	4.81	[99][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	33.58	1.46	30.91	1.31	12.59	1.13	this work
nitrogen	experimental		126.19	3.40	313.30	-	-	-	[95]
	TD	SAFT framework	128.36	3.65	292.69	1.72	7.55	6.58	[103][c]
		SAFT framework	126.15	3.68	301.21	0.04	8.28	3.86	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	196.52	6.23	322.96	55.73	83.41	3.08	this work
oxygen	experimental	-	154.58	5.04	436.10	-	-	-	[95]
	TD	SAFT framework	158.06	5.47	399.18	2.25	8.37	8.47	[103][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	174.18	5.68	384.01	12.68	12.73	11.94	this work
fluorine	experimental	-	144.41	5.17	592.86	-	-	-	[95]
	TD	SAFT framework	146.74	5.60	533.16	1.61	8.36	10.07	[103][c]
		SAFT framework	146.14	5.66	559.26	1.19	9.39	5.67	[87][c]
		SAFT framework	144.14	5.65	553.58	0.19	9.26	6.63	[107][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	116.11	4.25	510.61	19.60	17.80	13.87	this work
carbon monoxide	experimental	-	132.86	3.49	303.90	-	-	-	[95]
	TD	SAFT framework	134.94	3.73	283.48	1.57	6.87	6.72	[103][c]
		SAFT framework	132.90	3.79	296.40	0.03	8.53	2.47	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	147.81	4.46	309.79	11.25	27.70	1.94	this work
carbon dioxide	experimental	-	304.13	7.38	467.60	-	-	-	[95]
	TD	SAFT framework	308.31	8.41	502.66	1.37	14.02	7.50	[103][c]
		SAFT framework	303.84	8.06	471.76	0.10	9.24	0.89	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	344.69	9.50	446.31	13.34	28.82	4.55	this work
ethane	experimental	-	305.32	4.87	206.18	-	-	-	[95]
	TD	SAFT framework	312.34	5.39	190.62	2.30	10.65	7.55	[103][c]
		SAFT framework	310.99	5.47	205.39	1.86	12.27	0.38	[87][c]
		SAFT framework	305.54	5.41	203.68	0.07	11.09	1.21	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	331.99	6.06	201.86	8.73	24.44	2.10	this work
propane	experimental	-	369.89	4.25	220.00	-	-	-	[95]
	TD	SAFT framework	375.81	4.74	213.47	1.60	11.52	2.97	[103][c]
		SAFT framework	375.73	4.76	219.65	1.58	11.91	0.16	[87][c]
		SAFT framework	369.84	4.72	225.99	0.01	11.10	2.72	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	403.99	5.20	209.21	9.22	22.37	4.91	this work
n-butane	experimental	-	425.13	3.80	228.00	-	-	-	[95]
	TD	SAFT framework	432.77	4.34	239.17	1.80	14.28	4.90	[103][c]
		SAFT framework	432.10	4.26	227.52	1.64	12.09	0.21	[87][c]
		SAFT framework	424.89	4.17	223.69	0.06	9.81	1.89	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	475.85	4.79	216.86	11.93	26.19	4.89	this work
isobutane	experimental	-	407.81	3.63	225.50	-	-	-	[95]
	TD	SAFT framework	413.45	4.09	233.52	1.38	12.78	3.56	[103][c]
		SAFT framework	407.57	3.95	219.41	0.06	8.98	2.70	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	479.96	4.93	220.02	17.69	35.93	2.43	this work
ethene	experimental	-	282.35	5.04	214.20	-	-	-	[95]
	TD	SAFT framework	288.22	5.52	195.04	2.08	9.51	8.94	[103][c]
		SAFT framework	282.56	5.61	211.01	0.07	11.35	1.49	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	280.99	5.76	218.33	0.48	14.33	1.93	this work
propene	experimental	-	364.21	4.56	229.60	-	-	-	[95]
	TD	SAFT framework	371.43	5.05	210.96	1.98	10.97	8.12	[103][c]
		SAFT framework	364.25	5.10	235.50	0.01	12.04	2.57	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	456.61	7.66	259.31	25.37	68.10	12.94	this work
propyne	experimental	-	402.38	5.63	244.92	-	-	-	[95]
	TD	SAFT framework	414.25	6.38	225.32	2.95	13.38	8.00	[103][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	506.70	8.29	240.90	25.93	47.38	1.64	this work
cyclopropane	experimental	-	398.30	5.58	258.50	-	-	-	[95]
	TD	SAFT framework	408.48	6.33	240.31	2.56	13.38	7.04	[103][c]
		SAFT framework	398.85	6.38	263.99	0.14	14.32	2.12	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	469.20	7.47	246.27	17.80	33.95	4.73	this work
benzene	experimental	-	562.02	4.91	304.70	-	-	-	[95]
	TD	SAFT framework	573.39	5.47	277.98	2.02	11.50	8.77	[103][c]
		SAFT framework	567.76	5.52	308.19	1.02	12.41	1.15	[87][c]
		SAFT framework	561.54	5.46	304.17	0.09	11.20	0.17	[104][c]
	BU (S)	MP2/aug-cc-pVTZ (A)	1005.75	11.53	330.22	78.95	135.01	8.38	this work

[a]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods, and BU (P) refers to bottom-up methodology and perturbational methods. [b](A) indicates that approach A objective function was minimized in the parameter fitting procedure. [c]Percentage deviation values recalculated in this work.

Table D12 – AARD% values for saturation properties for fifteen molecules of different chemical classes using top-down and bottom-up-based Mie potential parameters in the SAFT-VR Mie equation of state.

			AARD% values		
Molecule	Methodology[b]	Context[c]	p_{sat}	ΔH_{vap}	Reference
hydrogen	TD	SAFT framework	27.39	33.67	[99][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	64.17	70.11	this work
nitrogen	TD	SAFT framework	0.40	2.16	[103][e]
		SAFT framework	7.71	2.91	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	95.81	100.31	this work
oxygen	TD	SAFT framework	0.58	1.78	[103][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	60.64	27.53	this work
fluorine	TD	SAFT framework	0.34	0.85	[103][e]
		SAFT framework	0.60	1.14	[87][e]
		SAFT framework	10.76	6.12	[107][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	821.72	36.66	this work
carbon monoxide	TD	SAFT framework	0.34	2.14	[103][e]
		SAFT framework	7.90	2.50	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	36.44	23.00	this work
carbon dioxide	TD	SAFT framework	0.45	5.14	[103][e]
		SAFT framework	8.51	1.07	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	31.90	16.41	this work
ethane	TD	SAFT framework	0.67	2.51	[103][e]
		SAFT framework	0.70	3.10	[87][e]
		SAFT framework	17.60	7.00	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	26.47	11.85	this work
propane	TD	SAFT framework	1.19	2.79	[103][e]
		SAFT framework	0.75	2.92	[87][e]
		SAFT framework	19.97	7.98	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	25.44	9.92	this work
n-butane	TD	SAFT framework	0.27	3.10	[103][e]
		SAFT framework	0.75	2.91	[87][e]
		SAFT framework	11.72	0.87	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	35.12	12.12	this work
isobutane	TD	SAFT framework	0.76	2.77	[103][e]
		SAFT framework	8.65	1.03	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	57.95	20.56	this work
ethene	TD	SAFT framework	0.38	2.30	[103][e]
		SAFT framework	12.12	4.93	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	16.26	4.46	this work
propene	TD	SAFT framework	1.07	2.57	[103][e]
		SAFT framework	17.60	7.44	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	74.59	36.71	this work
propyne	TD	SAFT framework	1.35	10.12	[103][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	66.99	59.39	this work
cyclopropane	TD	SAFT framework	0.38	6.51	[103][e]
		SAFT framework	15.00	4.08	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	53.05	40.24	this work
benzene	TD	SAFT framework	0.36	3.95	[103][e]
		SAFT framework	1.72	4.02	[87][e]
		SAFT framework	10.70	1.28	[104][e]
	BU (S)	MP2/aug-cc-pVTZ (A)	99.15	131.95	this work

[a]SAFT-VR Mie EoS refers to SAFT-VR Mie equation of state and MD simulations refers to molecular dynamics simulations. [b]TD refers to top-down methodology, BU (S) refers to bottom-up methodology and supermolecular methods. [c](A) indicates that approach A objective function was minimized in the parameter fitting procedure. [d]Parameters obtained through personal communication with professor Erich A. Müller (June 3rd, 2021), based on published works [106, 107]. [e]AARD% values calculated for the datasets of this work.

Table D13 – AARD% values for thermophysical properties of 26 molecules from different chemical classes obtained using a predictive SAFT-VR Mie equation of state with parameters derived from *ab initio* calculations performed in this study.

		AARD% values				
Chemical class	Molecule	ρ	c_v	c_p	c_{sound}	μ_{JT}
light gases	helium	1.51	4.20	5.11	1.79	9.65
	neon	6.70	1.20	4.84	3.07	109.30
	argon	2.92	1.87	2.37	2.29	67.52
	krypton	1.83	3.02	2.21	3.73	45.70
	hydrogen	1.21	7.89	6.97	1.41	17.83
	nitrogen	18.05	3.44	11.59	12.89	462.50
	oxygen	2.60	2.17	4.75	6.04	90.38
	fluorine	18.33	8.30	14.04	15.43	167.89
	carbon monoxide	5.75	1.62	2.50	2.88	29.57
	carbon dioxide	8.57	5.81	9.27	8.86	70.16
linear hydrocarbons	methane	1.46	1.45	2.62	1.00	-
	ethane	2.35	3.70	2.00	4.16	9.24
	propane	1.44	7.86	5.35	6.18	7.66
	n-butane	1.63	4.33	3.09	8.37	6.88
branched hydrocarbons	isobutane	3.14	4.14	3.20	7.83	6.27
unsatured hydrocarbons	ethene	1.85	8.06	11.06	12.02	253.40
	propene	21.41	2.78	1.44	10.98	23.14
	propyne	22.26	10.39	18.86	64.81	503.32
cyclic hydrocarbons	cyclopropane	13.52	6.59	11.63	51.52	536.41
aromatic hydrocarbons	benzene	28.26	6.47	3.95	62.88	21.50
refrigerants	tetrafluormethane	3.26	2.93	5.16	2.46	-
	fluoromethane	21.46	108.29	109.27	12.41	-
	chlorotrifluoromethane	17.88	188.50	214.85	22.96	-
	dichlorodifluoromethane	41.36	101.64	89.07	66.42	-
	chlorodifluoromethane	16.25	161.16	181.33	20.17	-
	dichlorofluoromethane	5.30	99.54	80.87	26.57	-

Table D14 – Percentage deviation in relation to the real critical point of 26 molecules from different chemical classes obtained using a predictive SAFT-VR Mie equation of state with parameters derived from *ab initio* calculations performed in this study.

		Deviation from NIST data (%)		
Chemical class	Molecule	T_{crit}	p_{crit}	ρ_{crit}
light gases	helium	105.45	176.87	20.95
	neon	47.80	55.23	22.92
	argon	1.11	5.00	6.32
	krypton	3.95	10.37	5.60
	hydrogen	1.31	12.59	1.13
	nitrogen	55.73	83.41	3.08
	oxygen	12.68	12.73	11.94
	fluorine	19.60	17.80	13.87
	carbon monoxide	11.25	27.70	1.94
	carbon dioxide	13.34	28.82	4.55
linear hydrocarbons	methane	0.77	9.42	4.65
	ethane	8.73	24.44	2.10
	propane	9.22	22.37	4.91
	n-butane	11.93	26.19	4.89
branched hydrocarbons	isobutane	17.69	35.93	2.43
unsatured hydrocarbons	ethene	0.48	14.33	1.93
	propene	25.37	68.10	12.94
	propyne	25.93	47.38	1.64
cyclic hydrocarbons	cyclopropane	17.80	33.95	4.73
aromatic hydrocarbons	benzene	78.95	135.01	8.38
refrigerants	tetrafluormethane	3.76	9.47	10.13
	fluoromethane	0.96	2.97	24.58
	chlorotrifluoromethane	12.11	18.10	10.92
	dichlorodifluoromethane	27.99	41.73	6.45
	chlorodifluoromethane	18.20	25.63	12.32
	dichlorofluoromethane	2.95	8.09	13.93

Table D15 – AARD% values for saturation properties of 26 molecules from different chemical classes obtained using a predictive SAFT-VR Mie equation of state with parameters derived from *ab initio* calculations performed in this study.

		AARD% values	
Chemical class	Molecule	p_{sat}	ΔH_{vap}
light gases	helium	-	-
	neon	-	-
	argon	15.31	1.75
	krypton	14.37	18.09
	hydrogen	64.17	70.11
	nitrogen	95.81	100.31
	oxygen	60.64	27.53
	fluorine	821.72	36.66
	carbon monoxide	36.44	23.00
	carbon dioxide	31.90	16.41
linear hydrocarbons	methane	28.39	4.23 -
	ethane	26.47	11.85
	propane	25.44	9.92
	n-butane	35.12	12.12
branched hydrocarbons	isobutane	57.95	20.56
unsatured hydrocarbons	ethene	16.26	4.46
	propene	74.59	36.71
	propyne	66.99	59.39
cyclic hydrocarbons	cyclopropane	53.05	40.24
aromatic hydrocarbons	benzene	99.15	131.95
refrigerants	tetrafluormethane	18.79	8.97 -
	fluoromethane	150.23	20.07 -
	chlorotrifluoromethane	33.24	9.19 -
	dichlorodifluoromethane	80.48	39.17 -
	chlorodifluoromethane	94.34	65.20 -
	dichlorofluoromethane	1163.40	42.81 -

APPENDIX E

Prediction of thermophysical properties for some substituted-methane compounds

Description: This Appendix presents figures referring to the prediction of thermophysical properties that were not presented for some molecules throughout the dissertation. On the following pages of this Appendix, predictions of thermophysical properties for fluoromethane, chlorotrifluoromethane, dichlorodifluoromethane, dichlorofluoromethane, and chlorodifluoromethane can be found.

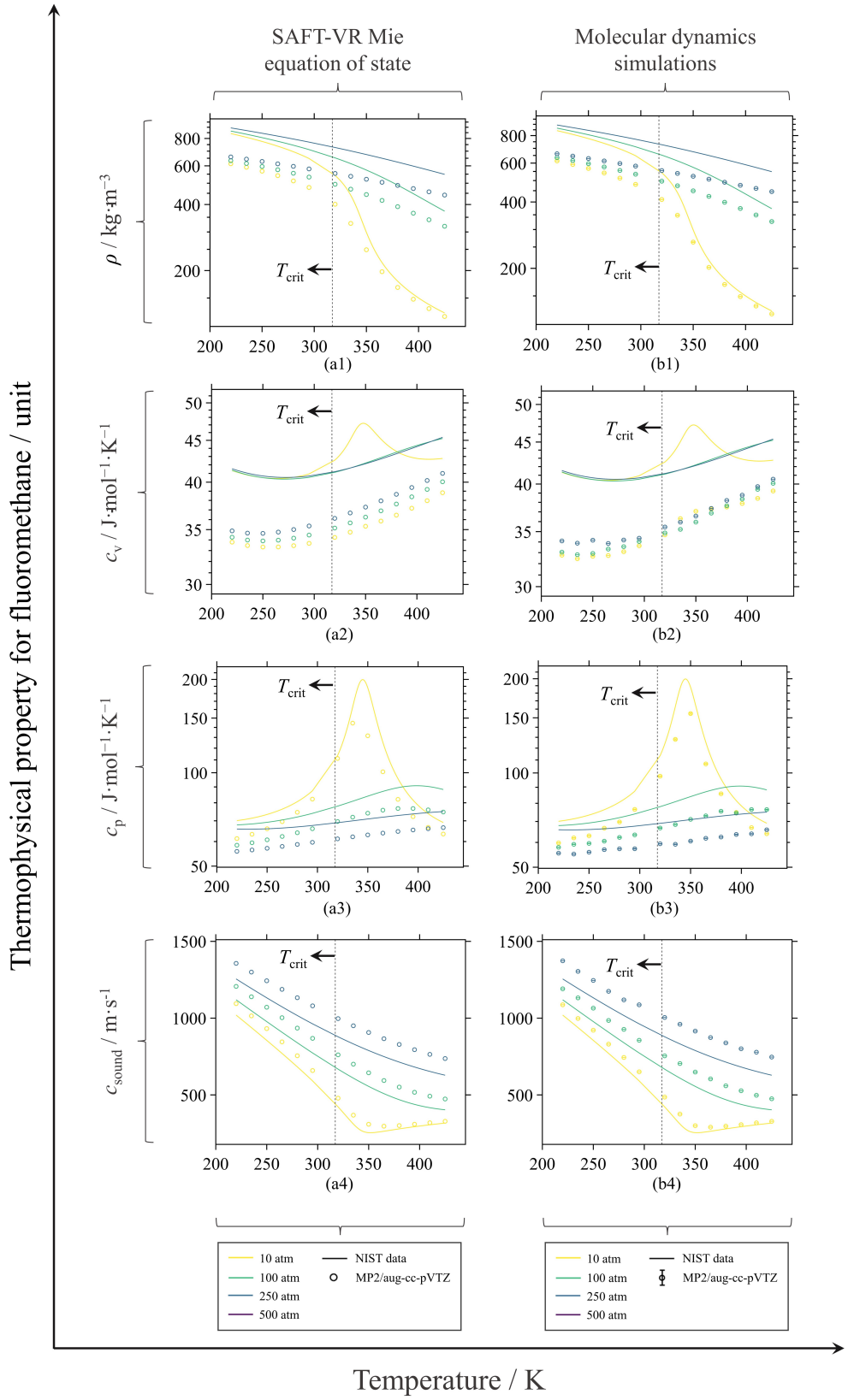

Figure E1 – Predicted thermophysical properties for fluoromethane using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Prediction of (a1) ρ, (b1) c_v, (c1) c_p, and (d1) c_{sound} by SAFT-VR Mie equation of state. Prediction of (a2) ρ, (b2) c_v, (c2) c_p, and (d2) c_{sound} by molecular dynamics simulations.

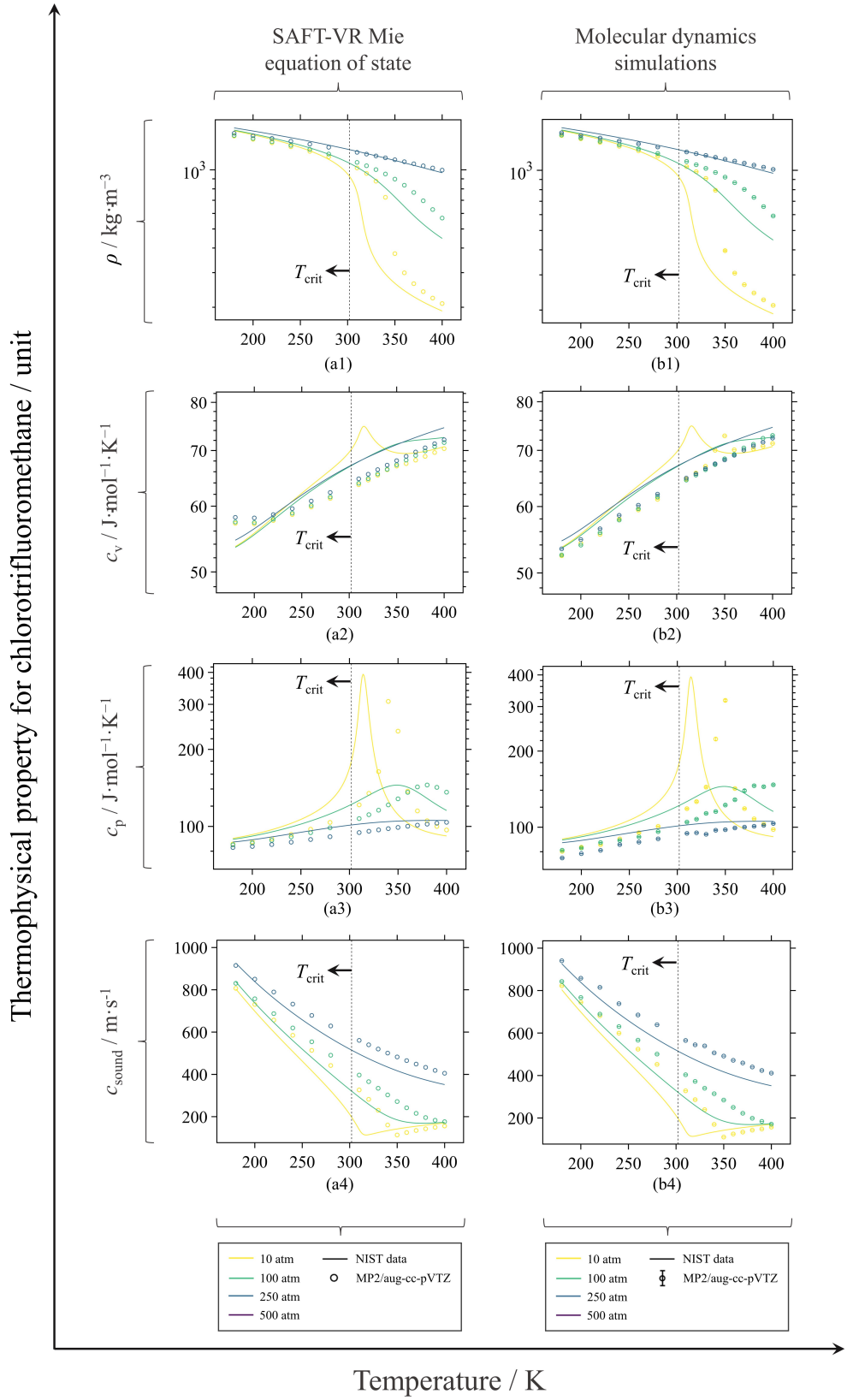

Figure E2 – Predicted thermophysical properties for chlorotrifluoromethane using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Prediction of (a1) ρ, (b1) c_v, (c1) c_p, and (d1) c_{sound} by SAFT-VR Mie equation of state. Prediction of (a2) ρ, (b2) c_v, (c2) c_p, and (d2) c_{sound} by molecular dynamics simulations.

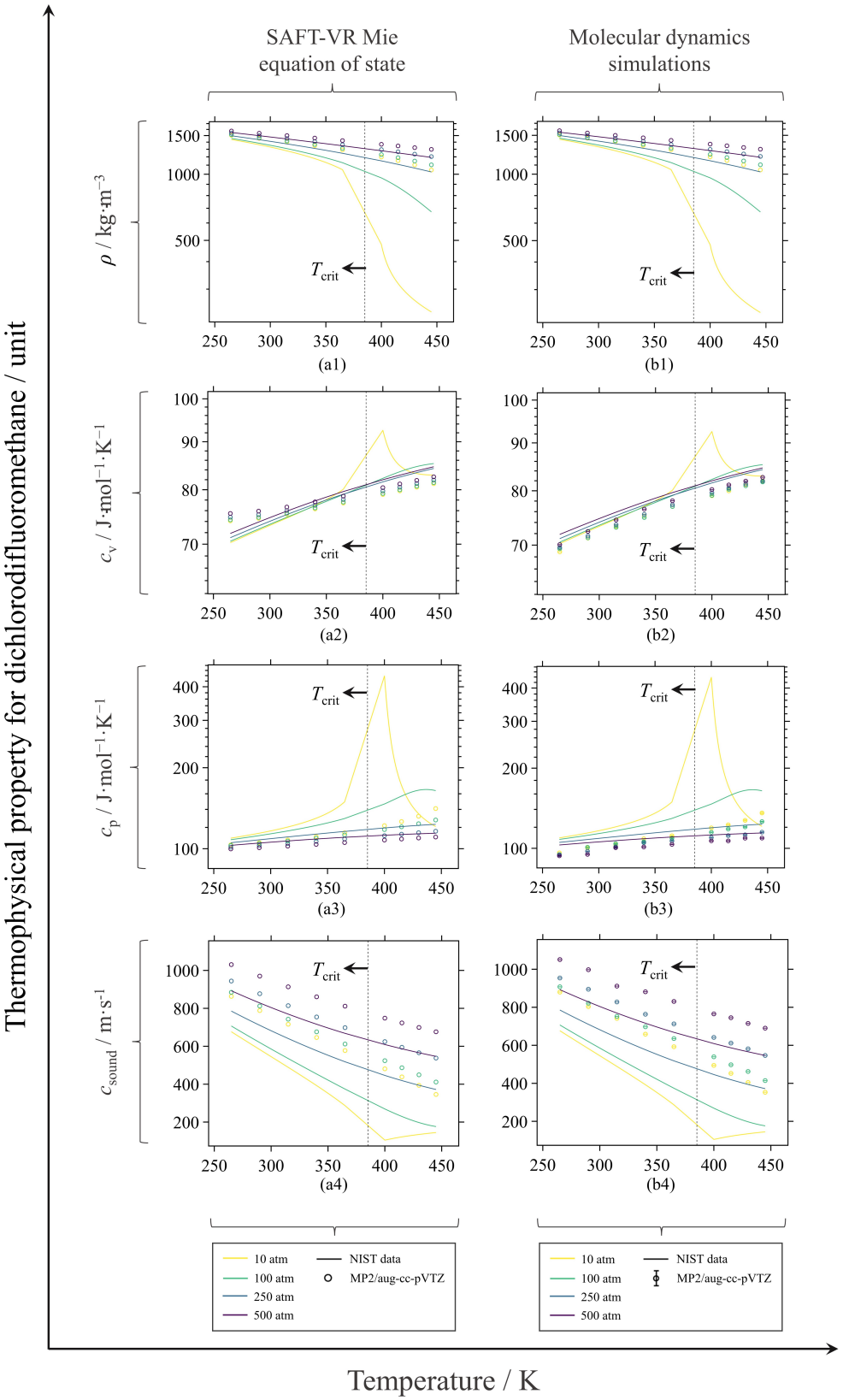

Figure E3 – Predicted thermophysical properties for dichlorodifluoromethane using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Prediction of (a1) ρ, (b1) c_v, (c1) c_p, and (d1) c_{sound} by SAFT-VR Mie equation of state. Prediction of (a2) ρ, (b2) c_v, (c2) c_p, and (d2) c_{sound} by molecular dynamics simulations.

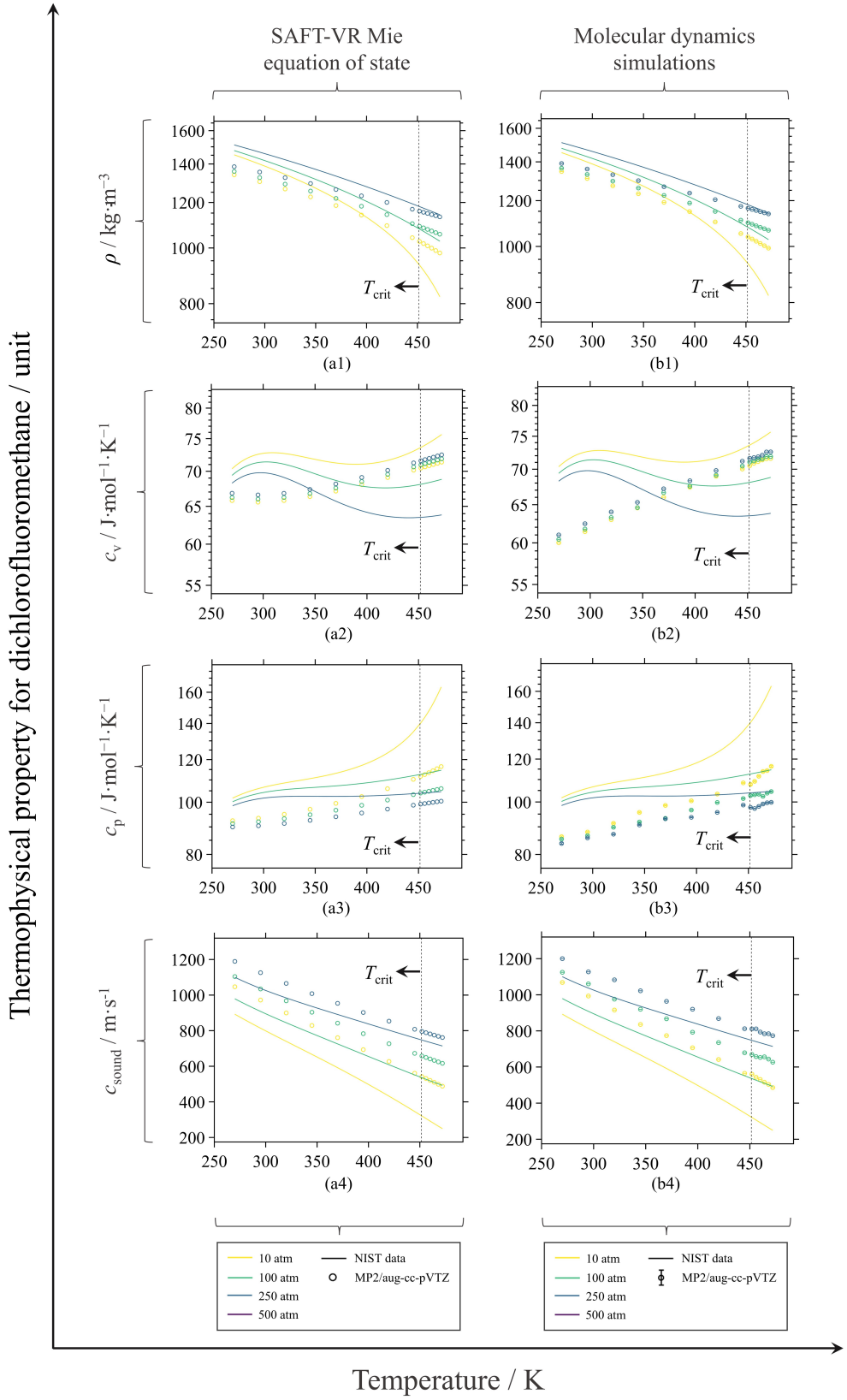

Figure E4 – Predicted thermophysical properties for dichlorofluoromethane using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Prediction of (a1) ρ, (b1) c_v, (c1) c_p, and (d1) c_{sound} by SAFT-VR Mie equation of state. Prediction of (a2) ρ, (b2) c_v, (c2) c_p, and (d2) c_{sound} by molecular dynamics simulations.

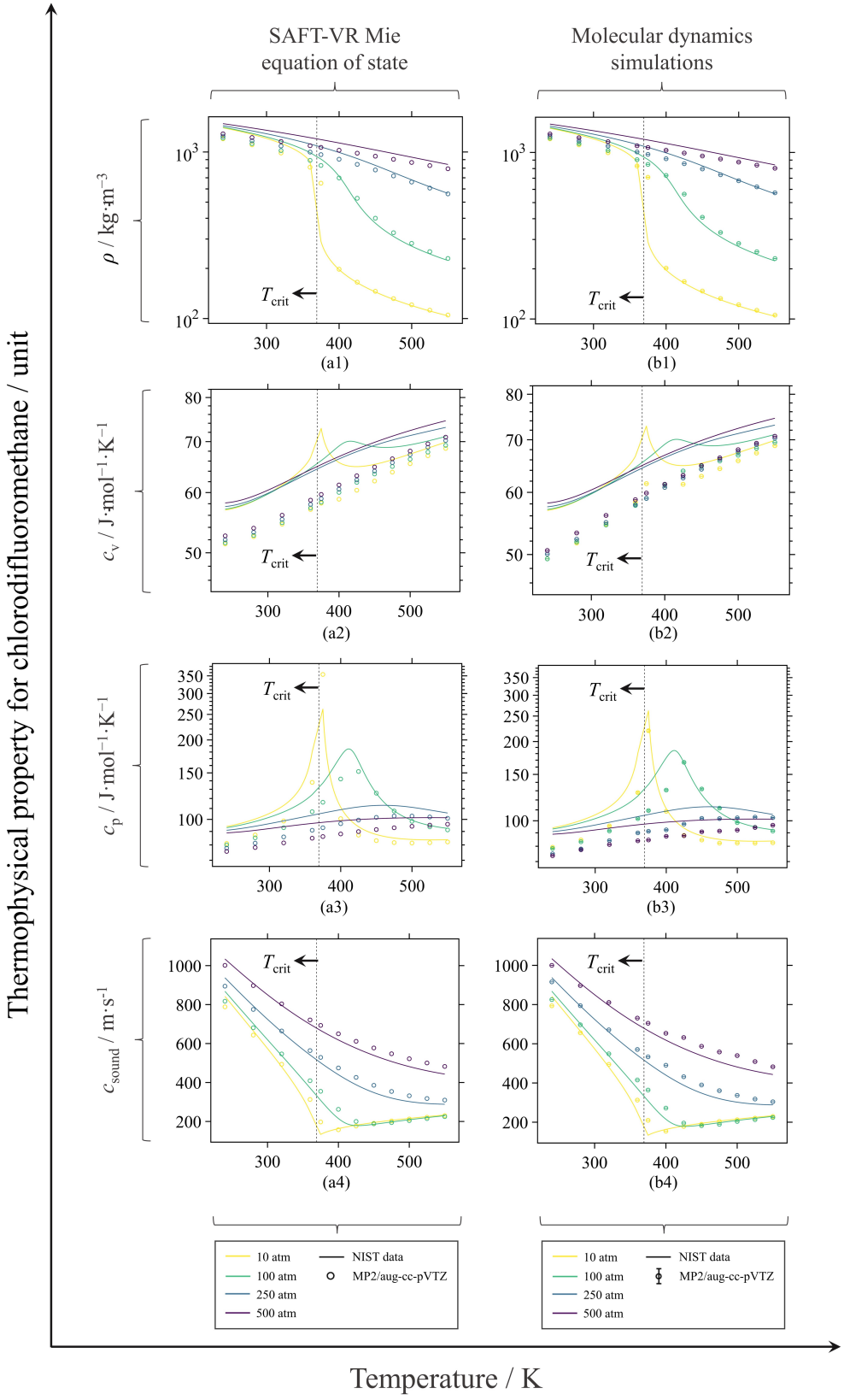

Figure E5 – Predicted thermophysical properties for chlorodifluoromethane using approach A bottom-up-based Mie potential parameters from MP2/aug-cc-pVTZ theory level energies in the SAFT-VR Mie equation of state. Prediction of (a1) ρ, (b1) c_v, (c1) c_p, and (d1) c_{sound} by SAFT-VR Mie equation of state. Prediction of (a2) ρ, (b2) c_v, (c2) c_p, and (d2) c_{sound} by molecular dynamics simulations.