



**HAL**  
open science

## Characterizing Molecular Interactions in Chemical Systems

David Guenther, Roberto Alvarez-Boto, Julia Contreras-Garcia, Jean-Philip Piquemal, Julien Tierny

► **To cite this version:**

David Guenther, Roberto Alvarez-Boto, Julia Contreras-Garcia, Jean-Philip Piquemal, Julien Tierny. Characterizing Molecular Interactions in Chemical Systems. *IEEE Transactions on Visualization and Computer Graphics*, 2014, 20 (12), pp.2476-2485. 10.1109/TVCG.2014.2346403 . hal-01146475

**HAL Id: hal-01146475**

**<https://hal.science/hal-01146475>**

Submitted on 1 Oct 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Characterizing Molecular Interactions in Chemical Systems

David Günther, Roberto A. Boto, Julia Contreras-Garcia, Jean-Philip Piquemal, Julien Tierny

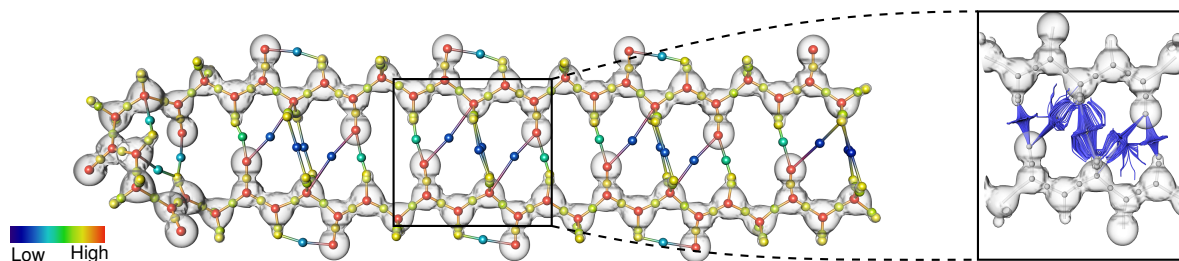


Fig. 1. Visual and quantitative exploration of covalent and noncovalent bonds in the  $\beta$ -sheet polipeptide. Our analysis enables to visualize, enumerate, classify, and investigate molecular interactions in complex chemical systems. In this example, the amplitude of the signed electron density ( $|\bar{\rho}|$ ), color-coded from blue to red) enables to distinguish covalent bonds (yellow) from hydrogen bonds (cyan) and van der Waals interactions (dark blue). While the numerical integration of  $\nabla\bar{\rho}$  (right inset) enables to visually distinguish the latter two types of interactions, our combinatorial pipeline robustly extracts these features to support further quantitative analysis. In particular, our algorithm reveals the repeating pattern (black frame) of noncovalent interactions responsible for the folding of this molecule, which decomposes it in unitary building blocks corresponding to the elementary amino-acids composing the molecule.

**Abstract**— Interactions between atoms have a major influence on the chemical properties of molecular systems. While covalent interactions impose the structural integrity of molecules, noncovalent interactions govern more subtle phenomena such as protein folding, bonding or self assembly. The understanding of these types of interactions is necessary for the interpretation of many biological processes and chemical design tasks. While traditionally the electron density is analyzed to interpret the quantum chemistry of a molecular system, noncovalent interactions are characterized by low electron densities and only slight variations of them – challenging their extraction and characterization. Recently, the signed electron density and the reduced gradient, two scalar fields derived from the electron density, have drawn much attention in quantum chemistry since they enable a qualitative visualization of these interactions even in complex molecular systems and experimental measurements. In this work, we present the first combinatorial algorithm for the automated extraction and characterization of covalent and noncovalent interactions in molecular systems. The proposed algorithm is based on a joint topological analysis of the signed electron density and the reduced gradient. Combining the connectivity information of the critical points of these two scalar fields enables to visualize, enumerate, classify and investigate molecular interactions in a robust manner. Experiments on a variety of molecular systems, from simple dimers to proteins or DNA, demonstrate the ability of our technique to robustly extract these interactions and to reveal their structural relations to the atoms and bonds forming the molecules. For simple systems, our analysis corroborates the observations made by the chemists while it provides new visual and quantitative insights on chemical interactions for larger molecular systems.

**Index Terms**—Molecular Chemistry, Topological Data Analysis, Morse-Smale Complex, Join Tree.

## 1 INTRODUCTION

The chemical properties of a molecular system are mainly governed by the interactions between the composing atoms. In particular, an interaction type of special interest is covalent bonding, which describes the sharing of electrons in between atoms. *Covalent bonds* have been widely studied since the early twentieth century [36]. They give rise to the chemical structure of a molecular system, and are deeply in-

vestigated in molecular chemistry [47]. When dealing with complex molecular systems, a second type of interactions governs many chemical phenomena: *noncovalent interactions*. These are responsible for the bonding between several molecules and the folding of single molecules onto themselves. Examples of chemical processes driven by such interactions include the bonding between a protein and a drug, a catalyst and its substrate, or self-assembly materials. Thus, the understanding of these interactions is necessary for the interpretation of many biological processes and chemical design tasks (e.g., pharmaceuticals, nano-technology).

In contrast to covalent bonds, noncovalent interactions release much less energy and are characterized by low electron density values and only slight value variations. This challenges their extraction and analysis by solely studying the electron density. Recently, the signed electron density and the reduced gradient have drawn much attention in the chemistry community. These scalar quantities are derived from the electron density and enable a qualitative visualization of the interactions [11, 30]. However, the analysis of these quantities is mainly done manually [11, 16, 30, 35]. An automated extraction and characterization of encoded chemical interactions (in terms of the involved atoms) is still an open issue.

In theoretical chemistry, the characterization of molecular interactions has been shown to relate to the topological structure – the critical points and their relation – of the electron density [1, 38]. Typically, covalent bonds given by the electron density are analyzed using nu-

- D. Günther is with Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, Paris, France. E-mail: gunther@telecom-paristech.fr.
- R. A. Boto is with Sorbonne Universités, UPMC Univ Paris 06, ICS, and Sorbonne Universités, UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005, Paris, France. E-mail: robalboto@gmail.com
- J. Contreras-Garcia, and J.P. Piquemal are with Sorbonne Universités, UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, and CNRS, UMR 7616, LCT, F-75005, Paris, France. E-mail: {julia.contreras,jpp}@lct.jussieu.fr
- J. Tierny is with CNRS LIP6, UPMC, Sorbonne Universités, CNRS LTCI, Télécom ParisTech, Paris, France. E-mail: julien.tierny@lip6.fr

Manuscript received 31 Mar. 2014; accepted 1 Aug. 2014; date of publication xx xxx 2014; date of current version xx xxx 2014.

For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

merical techniques. However, combinatorial counter-parts have been introduced in the visualization community [7, 8, 18, 25, 27, 53]. Benefiting from their robustness, they have been shown to capture covalent structures efficiently [25, 28]. In this work, we show how these kind of approaches can be generalized for the simultaneous extraction and characterization of covalent and noncovalent interactions in molecular systems. While the signed electron density and the reduced gradient have been shown to emphasize noncovalent interactions, these scalar fields challenge existing topological analysis techniques. In particular, the revealing of the noncovalent interactions comes at the price of a complex topological structure occluding the features of interest. In this paper, we show how a joint topological analysis of the two fields enables to derive an automated combinatorial algorithm for the extraction of covalent and noncovalent bonds and of the involved atoms. These bonds enable us to propose an algorithm which extracts combinatorially *repulsive bond-cycles*. This joint analysis reveals the atoms and bonds taking part in each of these interactions. We illustrate our method on a variety of molecular systems ranging from simple dimers to proteins or DNA. Our experiments demonstrate the ability of our technique to robustly extract these features and to reveal their structural relations to the atoms forming the molecules.

## 1.1 Related Work

Molecular data received a lot of attention in the visualization community in recent years [43]. The complexity of such data challenges rendering, exploration, and analysis tasks.

**Molecular Data Visualization** The traditional ball and stick representation of molecules, which describes the covalent bonds given by the Lewis dot diagram [36], challenges rendering techniques for macromolecular structures. Specific approaches exploiting GPU capabilities have been designed for this task [12]. An alternative visual representation of large molecular systems is the notion of molecular surfaces [15, 24, 49]. These are described by implicit formulations taking into account the position of the atoms, their van der Waals radii, and a probe radius. Several techniques have been proposed to efficiently render such implicit surfaces [2, 32, 45] or to generate illustrative or abstract visualizations [13, 54]. Moreover, several approaches have been proposed to emphasize certain features of molecular surfaces such as cavities [33, 46], pockets [34, 58] or paths to bonding sites [37]. Molecular surfaces have also been used as a driving representation for the geometrical analysis of molecular docking [4, 10, 42], to predict the preferred orientation of ligand bonds inside a protein. While molecular surfaces provide a convenient approximation of the shape of molecules for large systems, their implicit formulation only takes geometrical information into account. To accurately analyze covalent and noncovalent interactions, the quantum nature of these processes has to be taken into account.

**Computational Quantum Chemistry** In theoretical chemistry, density functional theory [29] is a popular quantum mechanical model. It enables to simulate electron systems and to evaluate in a probabilistic manner the electron population of molecular systems. Bader [1] proposed a theoretical model which relates electron interactions within molecules to the critical points of the the electron density [38]. This theory plays an important role in the explanation and understanding of many chemical observations. While it directly relates covalent bonds to certain saddles of the electron density, it has been shown to fail at capturing the complex nature of noncovalent interactions for molecules for which experimental evidence of such interactions is given [35]. Recently, the signed electron density and the reduced gradient [30] have drawn much attention in quantum chemistry. Specific isovalues determined by NCI [16, 30] – a diagram showing the dependency between the (signed) electron density and the reduced gradient – reveal noncovalent interaction sites in-between atoms among the extracted isosurfaces. However, the selection and identification of the atoms involved in each interaction is done manually. An automated extraction and characterization of chemical interactions based on the reduced gradient and the signed electron density has not been addressed so far.

**Topological Data Analysis** Combinatorial techniques for the topological analysis of scalar fields have been widely studied in the visualization community due to their ability to robustly capture the structure of the input scalar field into high-level topological abstractions, such as contour trees [8, 9], Reeb graphs [6, 18, 44, 53] or Morse-Smale complexes [25, 27]. Regarding molecular data analysis, Beketayev et al. [5] introduced for example a topological technique to analyze the energy related to the transition of molecules from a stable state to another. Only few topological techniques have been applied to the analysis of scalar fields representing molecular structures. Weber et al. [55] described a topology-driven volume rendering technique, which has been applied to the nucleon probability. Theisel et al. [52] presented a numerical technique which revealed the covalent bonds of molecules based on the analysis of the electrostatic field. Günther et al. [25] showed that a similar analysis could be performed with a combinatorial approach based on the Morse-Smale complex. Gyulassy et al. [28] also used the Morse-Smale complex to reveal covalent bonds from the electron density following Bader’s theory [1]. However, no topological analysis technique has yet been developed for the extraction and characterization of noncovalent interactions, which we also address in this paper.

## 1.2 Contributions

This paper makes the following new contributions:

1. Combinatorial extraction of covalent and noncovalent bonds;
2. Combinatorial extraction of noncovalent repulsive interactions;
3. Visual and quantitative exploration of molecular interactions.

## 2 PRELIMINARIES

In this section, we discuss the most important terms of molecular chemistry and scalar field topology, and describe the properties of our input which are relevant for the subsequent analysis.

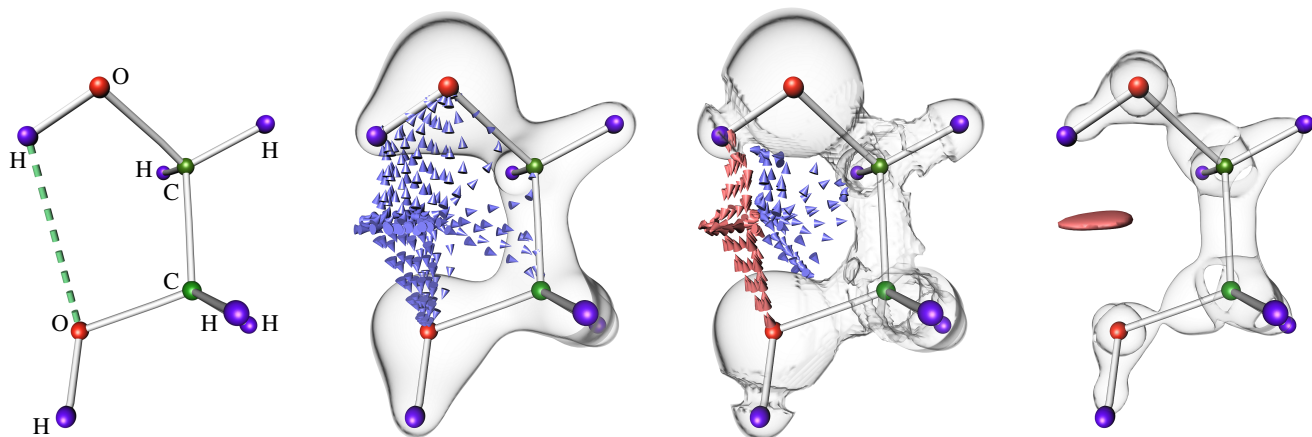
### 2.1 Interactions in Molecular Systems

Molecular interactions govern the structure of chemical systems by establishing attractive and repulsive balances in-between atoms. These interactions vary in strength and type. Here, we provide a brief characterization of them, and highlight some of their properties. We refer the reader to [47] for further details. For the purpose of our discussion, one mainly differentiates between two classes of chemical interactions:

A *covalent interaction* describes a chemical bond between atoms by sharing electrons. Based on electrostatic grounds, one can provide a simple picture of the physics behind it. Placing two atoms next to each-other, the two positively charged nuclei, i.e., the center of the atoms, both attract the outer negatively charged electrons. If the attraction is strong enough such that it overcomes the repulsion caused by the positively charged nuclei, a bond between the atoms is created. The bond represents an energetic equilibrium in which the electrons are equally attracted by the two nuclei. Thus, the two bonded atoms share these electrons. Multiple covalent interactions between atoms result in a system of bonded atoms called *molecule* – typically represented with balls (atoms) and sticks (covalent bonds) (Fig. 2(a)).

A *noncovalent interaction* does not involve the sharing of electrons. From an electrostatic point of view, they can be understood as weak electrostatic interactions between temporary and permanent partial charges. While this class of interaction constitutes the driving forces for intra-molecular folding and inter-molecular bonding, it spans a wide range of binding energies, which are typically from 1 to 15 kcal/mol and around one to two orders of magnitude smaller than covalent interactions. Hence, their extraction and characterization is much more involved. Noncovalent interactions can be caused by several physical phenomena and we detail here the most relevant ones:

- *Hydrogen bonds*: Among these electrostatic interactions, the hydrogen bonds deserve special attention due to their ever presence in biological systems. They link a hydrogen to an electronegative atom (e.g., oxygen, nitrogen, fluorine, carbon) and occur within



(a) Molecular structure of the 1,2-ethanediol molecule. Noncovalent bonding (green) occurs between the hydrogen and oxygen. (b) The  $\rho$ -field does not capture the noncovalent bond. The gradient flow  $\nabla\rho$  (blue arrows) uniformly covers all atoms. (c) The hydrogen-oxygen attraction is captured by  $\nabla\bar{\rho}$ . The attraction (blue arrows) closes a molecular cycle generating a repulsion (red arrows). (d) The noncovalent bond is given in  $s$  by an isolated component (red surface) highlighting the interaction site of the oxygen and hydrogen atom.

Fig. 2. Isosurfaces and gradient behavior (colored arrows) of the electron density  $\rho$ , its derived signed electron density  $\bar{\rho}$ , and the reduced gradient  $s$  for the 1,2-ethanediol molecule. Oxygen (O), carbon (C), and hydrogen (H) atoms are shown as red, green, and purple spheres, respectively. Covalent and noncovalent bonds are shown as white sticks and dashed green lines respectively.

the same molecule or in-between molecules. (dashed green line in Fig. 2(a)).

- *Van der Waals forces:* These attractive forces have a purely quantum mechanical nature. In particular, the constant movement of electrons around the nucleus transforms an atom into a fluctuating multipole. These temporary charges can cause attraction between close oppositely charged atoms yielding a stable bonding of weak energy. Although the force of an individual van der Waals bond is relatively weak, the cumulative effect of multiple of them may strongly influence the global structure of large molecular systems – as shown in many chemical reactions and protein-ligand interactions.
- *Steric repulsion:* These repulsive forces are short range interactions which occur when two atoms approach one another. Intuitively, they are due to the fact that too many electrons occupy the same space (Pauli principle). This can be pictured as forces occurring in regions of space bounded by negatively charged elements, such as covalent bonds and negatively charged atoms forming molecular cycles [23]. The localization of these interactions is of major importance for chemical design tasks since they indicate regions of space that cannot receive additional electrons.

## 2.2 Input Data

The input of our analysis are two scalar fields derived from the electron density: the signed electron density and the reduced gradient. In the following, we provide a brief description of these two fields. We refer the reader to [11, 14, 16, 30] for a more detailed description.

**Signed Electron Density.** In quantum chemistry, electrons behave simultaneously as waves and particles, which only allows for a probabilistic evaluation of their positions. The relative probability that electrons can be found in a particular location of a space  $\Omega$  is described by the *electron density*  $\rho : \Omega \rightarrow \mathbb{R}^+$ . Density cusps are expected at the nuclei, the center of the atoms, whereas charges decrease exponentially away from them. Thus, the nuclei dictate the overall behavior of  $\rho$ . Weak atomic interactions are very often occluded and cannot be directly computed or visualized. For instance, while the ethanediol molecule admits a noncovalent bond (dashed green line, Fig. 2(b)), this bond is not captured by the electron density  $\rho$  [35]. Investigating the flow of  $\nabla\rho$  in Fig. 2(b) reveals that the flow enters the molecular cycle from the outside and uniformly covers all atoms

forming the cycle. The circular structure shown in Fig. 2(a) is not captured by the flow while it is crucial for the analysis of attractive and repulsive interactions. A differentiation of these interactions solely based on the density  $\rho$  is not possible, in general. To compensate therefore, a direct investigation of the Hessian  $H\rho$  and its eigenvalues is needed [16]. Assuming the eigenvalues  $\lambda_i$  are given in increasing order, i.e.,  $\lambda_1 < \lambda_2 < \lambda_3$ , we observe the following behavior. In the vicinity of the nuclei all eigenvalues are negative. Away from it,  $\lambda_3$  becomes positive and varies along the internuclear axis representing covalent bonds.  $\lambda_1$  and  $\lambda_2$  describe the density variation orthogonal to this internuclear axis.  $\lambda_1$  represents the influence of the nuclei, and is always negative away from the nuclei. Contrarily,  $\lambda_2$  can be either positive or negative depending on the type of interaction. While attractive interactions concentrate electron charge perpendicular to the bond ( $\lambda_2 \leq 0$ ), repulsive interactions cause density depletion ( $\lambda_2 \geq 0$ ). Using this localized information, the *signed electron density*  $\bar{\rho}$  is defined as  $\bar{\rho} : \Omega \rightarrow \mathbb{R}$  with  $\bar{\rho}(x) = \text{sign}(\lambda_2(x))\rho(x)$  [16]. In contrast to  $\rho$  which only assesses the interaction strength of atoms, the signed electron density  $\bar{\rho}$  additionally enables the differentiation of attracting and repulsive interactions. Fig. 2(c) shows an isosurface of the signed electron density for the ethanediol molecule. In contrast to the electron density, the gradient  $\nabla\bar{\rho}$  captures nicely the attraction between the hydrogen and oxygen (red arrows), which forms a noncovalent bond creating a molecular cycle. This folded conformation also introduces repulsion in the molecule captured by  $\nabla\bar{\rho}$  (blue arrows).

While the electron density  $\rho$  is a continuous smooth scalar field away from the nuclei [21],  $\bar{\rho}$  is discontinuous, in general. While  $\bar{\rho}$  is necessary to investigate attractive and repulsive interactions, this property challenges its analysis. The theoretical setting of topological data analysis assumes continuous functions. To conform to this, we need to interpret  $\bar{\rho}$  as a result of a convolution with a mollifier [57]. However, given a mollification that interpolates the grid points, the convolution does not need to be performed explicitly in the discrete case, i.e., dealing with piecewise-linear or discrete Morse functions.

**Reduced Gradient.** To further reveal weak noncovalent interactions, the *reduced gradient*  $s : \Omega \rightarrow \mathbb{R}^+$  of  $\rho$  was introduced [30]

$$s = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}}. \quad (1)$$

The reduced density gradient  $s$  describes the deviation in atomic densities due to interactions [30]. Intuitively, covalent and noncovalent

interactions both appear in  $s$ . In the presence of such interactions,  $s$  reports a strong change in its values in regions of space between interacting atoms. The denominator of  $s$  reduces herein the influence of the nuclei due to their high electron densities. In contrast to the electron density,  $s$  shows large values in regions far from the nuclei. The electron density exponentially decreases towards zero, and  $\rho^{4/3}$  converges faster to zero than  $|\nabla\rho|$ . Thus,  $s$  behaves divergent at infinity [22]. Points at which  $\nabla\rho$  vanishes become zeros in  $s$  and are minima of  $s$ . In contrast to the infinite-behavior, the gradient  $\nabla\rho$  dominates the denominator  $\rho^{4/3}$  at those points.

The reduced gradient provides a qualitative visualization of chemical interactions by considering its isosurfaces chosen at appropriate values [11, 30]. Fig. 2(d) shows an isosurface of  $s$  for the ethane-diol molecule. The interaction of the hydrogen and oxygen causes the creation of a component of isosurface (highlighted in red). The investigation of such components, which describe regions of space where interactions occur, is the starting point of our approach. In particular, the topological analysis of  $s$  reveals *interaction sites* which enables us to focus on locations of space which are relevant from a chemical perspective.

### 2.3 Scalar Field Topology

In this section, we introduce the most relevant notions of scalar field topology. We refer the reader to [19] for a more comprehensive introduction into this topic. The topological structure of a scalar function  $f: \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$  is given by its critical points and their global relation. For the following discussion, we assume that  $f$  is a Morse function [41] and behaves transversal [51].

**Critical Points.** Points of the domain  $\Omega$  where the gradient of  $f$  vanishes are called *critical points*. These can be classified according to their index (i.e. the number of negative eigenvalues  $\lambda_i$  of the Hessian  $H$  of  $f$ ). In 3D, four types of critical points may occur: minima, 1-saddles, 2-saddles and maxima with indices 0, 1, 2, and 3 respectively. Morse theory [39] relates the critical points of  $f$  to the topological evolution (i.e. the creation or destruction of connected components, tunnels and voids) of its sub-level sets. For example, a connected component is born at a minimum and dies at a maximum. The merge and split of connected components occur at 1-saddles and 2-saddles, respectively. Similarly, the creation and destruction of tunnels also happen at 1-saddles and 2-saddles.

**Join Trees.** The sub-level set  $L^-(i)$  of a function  $f$  is defined by the set of points below a certain isovalue  $i$ :  $L^-(i) = \{p \in \Omega \mid f(p) \leq i\}$ . The evolution of the connected components of  $L^-(i)$  is captured by the *join tree*. This tree is an abstract graph capturing the merges of connected components of  $L^-(i)$  occurring at different isovalues. Each leaf in the tree represents a minimum and each interior node a 1-saddle. The branches connecting two nodes in the tree indicate which components merge at which isovalues. These branches can also be *augmented* [8] by considering the list of vertices of the sampled domain that are swept by the corresponding sub-level set component as the isovalue  $i$  increases.

**Morse-Smale Complex.** The join tree captures nicely the merging of sub-level set components. However, it does not represent isosurface genus changes. To represent these changes, the gradient  $\nabla f$  needs to be investigated. At minima and maxima, the gradient flow either emanates or terminates. At saddle points, however, we observe a mixed flow behavior. Given a 1-saddle, there are always two inflow directions entering the saddle, and a 2D manifold on which the flow is leaving the saddle. The two inflow directions are represented by integral lines of  $\nabla f$  – called *separatrices* – and connect the 1-saddle to at most two minima. The outflow manifold describes a 2D-separation between the flow emanating at the two minima. The boundary of the separation is given by separatrices linking 2-saddles and maxima. An analogue observation can be done for the separation of the flow which terminates at maxima. The connectivity between critical points based on the gradient flow gives rise to an abstract graph representation called the *Morse-Smale complex* [40, 19], which decomposes the domain  $\Omega$  into compartments where  $f$  behaves monotonically [39]. Note

that the information about the evolution of connected components of sub-level sets (captured by the join tree) is not directly accessible in the Morse-Smale complex.

**Computation.** The Morse-Smale complex can be computed numerically [56] but such a computation is prone to a variety of parameters which strongly influence the stability of the output. Combinatorial methods became recently successful in the computation of the join tree [8] and the Morse-Smale complex [26, 27, 50]. Their avoidance of explicit derivative evaluations makes the computations not only robust but also guarantees the consistency of the computed topological structure. In this work, we make use of existing combinatorial algorithms to compute the join tree [8] and the Morse-Smale complex [26, 48].

**Simplification.** In practice, the scalar function  $f$  is given on a discrete sampling of the domain  $\Omega$  such as a regular grid. Due to this, spurious topological structures may be introduced in the join tree and the Morse-Smale complex which occlude the essential structures of interest. To analyze efficiently the function  $f$ , those abstract graph representations need to be simplified. In recent years, simplification guided by topological persistence [20] became popular. This simplification allows the differentiation between small- and large-scale topological structure, and removes iteratively pairs of critical points from the abstract graphs guided by their persistence. While the join tree can be perfectly simplified, i.e., all non-essential minimum-saddle pairs can be removed, it is no longer the case for the Morse-Smale complex. In particular, the sequential simplification of the Morse-Smale complex to a prescribed number of critical points is an NP-complete problem [3, 31]. In practice, this results in saddle points which are spurious but not removable from the Morse-Smale complex.

In this work, we use this simplification strategy to reduce the topological complexity of the join tree of  $s$  and the Morse-Smale complex of  $\tilde{\rho}$ . We follow here the work [9, 25]. To circumvent the simplification problem in the Morse-Smale complex, we make use of the topological information of the reduced gradient  $s$ . This additional information enables us to remove all critical points which do not represent bonds from a chemical perspective.

## 3 TOPOLOGICAL REPRESENTATION OF INTERACTIONS

In this section, we describe a topological representation of molecular interactions that will be used in the next section to derive our characterization pipeline.

### 3.1 Feature Definition

Bader [1] proposed a theoretical model which relates molecular interactions to the critical points of the electron density [38]. In particular, nuclei can be considered as local maxima of the electron density  $\rho$  whereas bond interactions between atoms are represented by separatrices emanating from 2-saddles and ending in two distinct atoms. Fig. 3(a) shows these features extracted from the Morse-Smale complex of  $\rho$  on the ethane-diol. However, as observed by Lane et al. [35], even without any topological simplification,  $\rho$  exhibits no 2-saddle in between the upper-left hydrogen and the lower oxygen atoms (cf. dashed line in Fig. 2(a)). Thus, no separatrix connecting these two atoms can be extracted and no chemical bond is identified between them. However, as discussed by Lane et al. [35], it is established from experimental evidence that this molecule exhibits a noncovalent hydrogen bond connecting these two atoms (Fig. 2(a)). Since  $\tilde{\rho}$  and  $s$  better emphasize noncovalent interactions, we can overcome this limitation by transposing Bader’s model from  $\rho$  to the analysis of  $\tilde{\rho}$  and  $s$ . In particular, we focus on the following features.

**Atoms** Near the atoms, the flow of  $\nabla\rho$  exhibits an attractive behavior characterized by  $\lambda_2 < 0$  (Sec. 2.2). Thus,  $\tilde{\rho}$  is negative in the vicinity of the atoms. Moreover, since the electron population is maximal at the nuclei,  $|\tilde{\rho}|$  is also maximal in these locations. Thus atoms are represented in  $\tilde{\rho}$  by local minima. With regard to  $s$ , these atoms are also represented by minima of  $s$  since  $\nabla\rho$  vanishes in the atoms. For typical quantum chemical calculations with Gaussian bases, atoms are characterized by  $\tilde{\rho} < -0.35$  and  $s = 0$ .

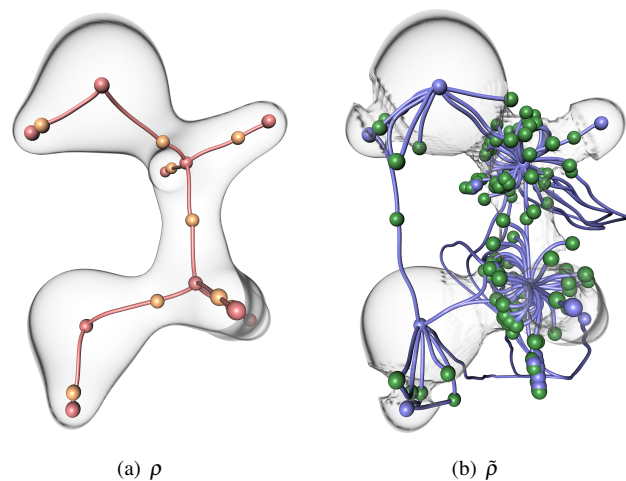


Fig. 3. Critical points (spheres) and separatrices (lines) in the ethanediol molecule for the considered scalar fields. Red, orange, green and blue spheres represent maxima, 2-saddles, 1-saddles and minima respectively. Red and blue lines represent separatrices connecting a 2-saddle to maxima and a 1-saddle to minima respectively.

**Bond Paths** By transposing Bader’s model to the analysis of  $\tilde{\rho}$ , a bonding interaction between two atoms is present if there exists a single line connecting the atoms and which is everywhere tangential to  $\nabla\tilde{\rho}$ . Since atoms are represented by minima of  $\tilde{\rho}$ , such a line is represented by two separatrices emanating from minima. Moreover, the existence of such a line between two minima implies the existence of a 1-saddle of  $\tilde{\rho}$ , which connects the two separatrices emanating from the minima. The latter saddle is called a *bond critical point*. Note that this saddle can be a merging or genus-change 1-saddle.

**Bonding Graph** The set of atoms and bond paths in a molecular system form the *Bonding Graph*. Each node of this graph represents an atom and its edges represent all attractive interactions in the system.

**Repulsive Bond Cycles** Bonds can form cycles in the bonding graph yielding steric repulsion (Sec. 2.1). Thus, we define a cycle of minimal length in the bonding graph as a repulsive bond cycle.

Fig. 3(b) shows the extraction of the minima and the ascending separatrices ending in 1-saddles, obtained from the Morse-Smale complex of  $\tilde{\rho}$ . In contrast to the topological analysis of  $\rho$  (Fig. 3(a)), this analysis reveals a bond path connecting the upper-left hydrogen and the lower oxygen atoms, hence, corroborating experimental observations [35]. Thus, this result further motivates the topological analysis of  $\tilde{\rho}$  for noncovalent bond extraction. Note that  $s$  also captures this interaction site by a component of its isosurface (Fig. 2(d)).

### 3.2 Technical Challenges

Although the Morse-Smale complex of  $\tilde{\rho}$  nicely captures covalent and noncovalent bonds, this comes with a price to pay in terms of topological complexity (Fig. 3(b)). This complexity occludes the features of interest and challenges their direct extraction, as illustrated in Fig. 4.

According to Sec. 3.1, atoms can be identified by considering minima of  $\tilde{\rho}$ . Bond paths can be identified by extracting from the Morse-Smale complex of  $\tilde{\rho}$  the two descending separatrices emanating from each 1-saddle. A bond path is considered *valid* if it connects two distinct atoms. Note that the robust extraction of the bond paths is of major importance for the extraction of the remaining features (covalent bonds, noncovalent bonds, Bonding Graph, repulsive bond cycles). However these features are also the most challenging to isolate, for the following reasons:

1. *Low persistence structures*: Spurious critical points are present in the Morse-Smale complex due to the sampled representation of  $\tilde{\rho}$ . While spurious 1-saddles can yield false bond paths, spurious minima can additionally yield interrupted bond

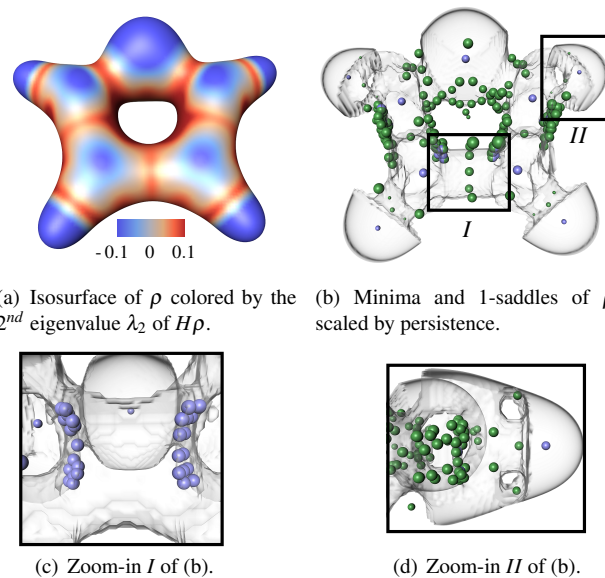


Fig. 4. Technical challenges associated with the extraction of molecular interactions from  $\tilde{\rho}$  on the furan molecule: beyond spurious low-persistence critical point pairs,  $\tilde{\rho}$  also exhibits high-persistence minimum-saddle (c) and saddle-saddle pairs (d) which occlude the critical points of interests (b) representing atoms and bond critical points.

paths by preventing descending separatrices emanating from a valid bond critical point to reach the corresponding atoms. Thus, the Morse-Smale complex needs to undergo topological simplification to (i) remove low-persistence minima and to (ii) connect bond critical points to atoms.

2. *Non-atom minima*: Additionally, at the edge of attractive ( $\tilde{\rho} < 0$ ) and repulsive ( $\tilde{\rho} > 0$ ) regions (Fig. 4(a)),  $\tilde{\rho}$  exhibits high-persistence minima, as illustrated in Fig. 4(c). These minima also need to be discarded to avoid interrupted bond paths as well as false positives in atom identification.
3. *Non-bonding saddles*: As illustrated in Fig. 4(d),  $\tilde{\rho}$  also includes high-persistence (genus-change) saddle-saddle critical point pairs. However, as discussed in Sec. 2.3, this category of critical point pair may not be removable through Morse-Smale complex simplification [31]. Thus, the removal of these outliers challenges existing topological techniques and motivates the joint analysis of the reduced gradient  $s$ .

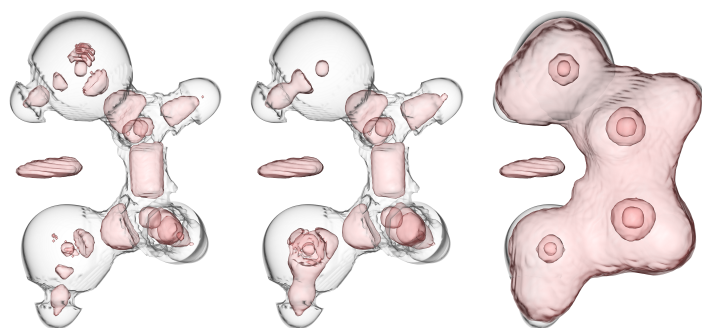
The above challenges motivate a tailored topological analysis pipeline described in the following section.

## 4 EXTRACTION OF MOLECULAR INTERACTIONS

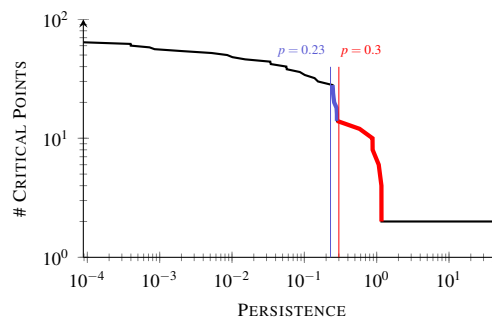
Based on the definitions and challenges introduced in the previous section, we now discuss how molecular interactions can be extracted. The extraction consists of three steps. Firstly, we discuss how the set of 1-saddles can be reduced to a set of possible bond-candidates. Given this candidate set, we present an algorithm to compute the Bonding Graph in the second step. This graph represents covalent and noncovalent bonds. Finally, we make use of the structure of the Bonding Graph to extract molecular repulsive interactions in the third step.

### 4.1 Interaction Sites

To address the third challenge of Sec. 3.2, we describe a pre-processing stage whose purpose is to isolate regions of space containing 1-saddles of  $\tilde{\rho}$  which are relevant bond critical point candidates from a chemical perspective. Interaction sites can be visualized by considering the regions of space bounded by the connected components of relevant isosurfaces of  $s$  [11, 30] (Sec. 2.2). However, each



(a) Interaction sites obtained from the sub-level set components of the reduced gradient  $s$  for three simplification levels: no simplification (left, 32 sites),  $p = 0.23$  (middle, 14 sites),  $p = 0.3$  (right, 6 sites).



(b) Persistence curve highlighted with two different simplification levels (blue and red vertical lines).

Fig. 5. Illustration of different simplification levels of the reduced gradient  $s$  based on the persistence  $p$  of critical point pairs. Reading the persistence curve from the right to the left side reveals a first slope (red) at the end of which interaction sites corresponding to covalent bonds have been simplified. The end of the next slope on the left ( $p = 0.23$ , blue) corresponds to a simplification level where both covalent and noncovalent interaction sites are maintained.

type of interaction is visualized by a specific isovalue. The birth of these components happens at early isovalues for covalent bonds. Contrarily, for noncovalent bonds, the birth happens at arbitrarily larger isovalues. To extract all interaction sites in a single procedure, we analyze the augmented Join Tree of  $s$  (Sec. 2.3).

The reduced gradient  $s$  exhibits low values in the vicinity of the nuclei and larger values away from it (Section 2.2). Moreover, as the isovalue increases, new connected components of isosurfaces appear in regions of spaces where interactions (covalent and noncovalent bonds) occur. The appearance of such components are characterized by minima of  $s$ . As the isovalue continues on increasing, the merge of these components occur at 1-saddles of  $s$ . As described in Sec. 2.3, these events are captured by the Join Tree of  $s$ . In particular, each branch of the Join Tree which contains a leaf (a minimum) represents a connected region of space containing a single minimum of  $s$ . Therefore, we extract each interaction site by considering each branch of the augmented Join Tree which contains a minimum.

Fig. 5(a) (left) shows the interaction sites (bounded by red surfaces) extracted from the analysis of the Join Tree of  $s$ . Due to the sampled representation of  $s$  and its numerical evaluation, spurious features are present – in particular in the vicinity of the atoms. To discard these, we apply a persistence based simplification of the augmented Join Tree. To select an appropriate simplification level, we manually inspect the persistence curve of the saddle-minimum pairs tracked by the Join-Tree (Fig. 5(b)) for the ethanediol molecule. This curve plots the evolution of the number of maintained critical point pairs for increasing persistence thresholds. Reading the curve from the right side (high persistence pairs) to the left (low persistence pairs) reveals a first slope (red in Fig. 5(b)) finishing with a sharp kink at  $p = 0.3$  (red vertical line). This level corresponds to the removal of the minima representing covalent bonds. Thus, at this level of simplification, the interaction sites extracted from the augmented Join Tree (Fig. 5(a), right) only reveal: the prominent atoms (oxygen and carbons), a larger component which represents the sub-level set containing all covalent interactions, and an isolated component representing a noncovalent interaction. The following slope to the left (blue) also finishes with a sharp kink at  $p = 0.23$  (vertical blue line). This point corresponds to a simplification level where the components representing covalent bonds are maintained (Fig. 5(a), middle). This observation was confirmed in all of our experiments. Thus, we select as an appropriate simplification level the end of the second slope (from the right side) on the persistence curve. As illustrated in Fig. 5(a) (middle), each interaction (covalent and noncovalent) is represented by a specific interaction site. Note that dominant atoms (oxygen and carbons) are also represented. However, the latter regions typically include no 1-saddle of  $\tilde{\rho}$  which enables to discard them in a straightforward manner. Note that at a technical level, in contrast to the Morse-Smale complex of  $\tilde{\rho}$ , all

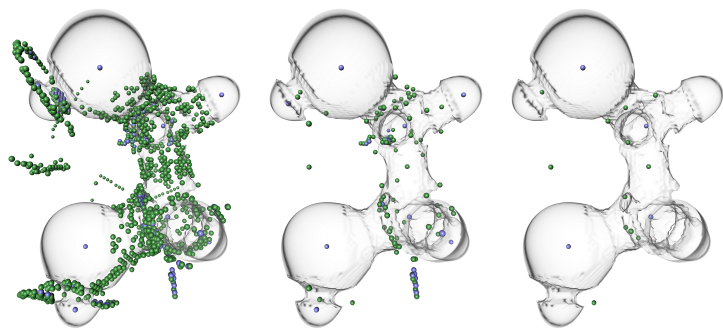
saddle-minimum pairs captured by the Join Tree of  $s$  can be simplified, which enables to overcome the technical challenge introduced by the incomplete simplification of the Morse-Smale complex of  $\tilde{\rho}$ .

## 4.2 Bonding Graph

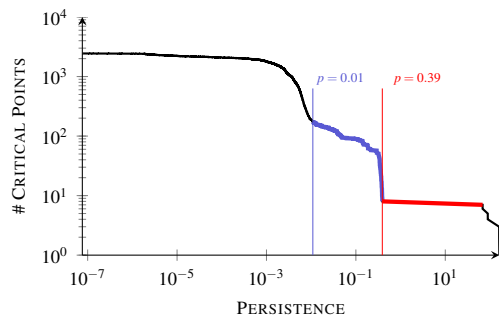
As discussed in Sec. 3.1, covalent and noncovalent bonds form a subset of the Morse-Smale complex. In the following, we show how the Morse-Smale complex of  $\tilde{\rho}$  can be reduced to the Bonding Graph representing all atoms and their bonds only.

1. **Removal of Low Persistence Structures.** To address the first challenge of Sec. 3.2, the Morse-Smale complex of  $\tilde{\rho}$  needs to be simplified. To find a representative simplification level, we analyze the persistence  $p$  of all minima and saddles. Fig. 6(b) shows such a curve for the ethanediol molecule in logarithmic scaling. The persistence curve exhibits a characteristic behavior. We can make use of this to guide the manual selection of an appropriate simplification level. Around 90% of the minima and saddles have a very low persistence ( $p < 0.01$ ). After the first drop down in this curve, two different persistence-ranges are apparent. In the first range ( $0.01 \leq p \leq 0.39$ ), the curve shows an almost constant slope (blue) indicating the transition from low to high persistent features. Fig. 6(a) (middle) shows the minima and 1-saddles at the simplification level  $p = 0.01$ . Most of the spurious structures are removed at this level, and the critical points representing atoms and bonds are revealed. However, there exists also minima and 1-saddles which do not represent molecular features. Increasing the persistence threshold now iteratively removes pairs of minima and 1-saddles. At the beginning of the second range ( $p = 0.39$ , red), only the minima representing the heavy carbon and oxygen atoms are still present. The minima representing hydrogens are already removed due to the small mass of hydrogens. However, 1-saddles which do not represent bonds are still visible in the interface between attractive and repulsive regions. We made the above observation in all of our experiments. To guarantee that all atoms and their even weak bonds are well represented, we select a simplification level after the first drop down in the persistence curve. However, the point of first drop down varies depending on the overall structure and energy of the molecular system.

2. **Removal of Non-Atom Minima.** As discussed in Sec. 3.2 (second challenge), there are minima which do not represent atoms. These are characterized by their location at the interface between the attractive and repulsive regions. Thus, their signed density value needs to be larger than those of the atoms. This allows a differentiation of the minima, and we select those with  $\tilde{\rho} > -0.35$ . Since those minima can interrupt bond paths, we need to remove them from the Morse-Smale complex yielding a new connectivity between the remaining minima and 1-saddles. We follow here the general idea of simplification as presented in [25, 28] but restricted to the selected set of minima. For



(a) Minima (blue) and 1-saddles (green) for three simplification levels: no simplification (left),  $p = 0.01$  (middle),  $p = 0.39$  (right).



(b) Persistence curve highlighted with two different simplification levels (blue and red line).

Fig. 6. Illustration of different simplification levels of the signed electron density  $\tilde{\rho}$  based on the persistence  $p$  of critical points. In the initial level minima and 1-saddles representing atoms and bonds are occluded by spurious critical points. Simplifying the Morse-Smale complex removes spurious critical point revealing the features of interest. However, not all minima and 1-saddles represent atoms and bonds even in the simplified Morse-Smale complex.

each minimum, we determine the lowest saddle in its neighborhood; and put all minima-saddle pairs with their weight, i.e., the height difference between minimum and saddle, in a priority queue. This queue is processed in ascending order starting with the pair with the lowest weight. This pair is removed from the Morse-Smale complex, which requires a subsequent update of the connectivity information of the critical points in the Morse-Smale complex. Due to the new connectivity, the weights of the minimum-saddle pairs also need to be updated yielding a new lowest minimum-saddle pair. We iteratively remove the selected minima until all of them are processed. This yields a Morse-Smale complex in which all minima represent atoms only.

**3. Removal of Non-Bonding Saddles.** To address the third challenge of Sec. 3.2, only the 1-saddles of  $\tilde{\rho}$  which are relevant from a chemical perspective should be considered for bond path extraction. Thus, we restrict the remainder of the analysis to the 1-saddles located in each of the interaction sites extracted in Sec. 4.1.

**4. Bonding Graph Extraction.** Each 1-saddle is connected to either one or two minima. Since bonding only occurs between atoms, we neglect all 1-saddles which are twice connected to the same minimum. From the remaining saddles, we collect the 1-saddles  $S^k$  and their connected minima  $M_1^k$  and  $M_2^k$  giving rise to a set of triplets  $(M_1^k, M_2^k, S^k)$ . It may happen that they exist two saddles  $S^{k_1}$  and  $S^{k_2}$  which end in the same pair of minima, i.e.,  $M_1^{k_1} = M_1^{k_2}$  and  $M_2^{k_1} = M_2^{k_2}$ . However, only one bond can exist between the same pair of atoms. Considering an evolution of the isovalue in  $\tilde{\rho}$ , bonding critical points are given by the earliest contact of the sub-level set components emanating from the two atoms. Hence, for each interaction site, we choose the 1-saddle  $S$  which minimizes its  $\tilde{\rho}$ -value as the bonding saddle representing the bond between  $M_1$  and  $M_2$ . Applying this procedure to all saddles  $S^k$  yields the Bonding Graph  $G = (N, E)$  in which all minima represent atoms and each pair of minima is only connected by a single 1-saddle representing their bonding. The nodes  $N$  of the undirected graph  $G$  are given by the minima and bonding 1-saddles, its edges  $E$  by the separatrices connecting the saddles to the minima.

Fig. 7 shows the different steps of our pipeline on the ethanediol molecule and illustrates its ability to isolate the features of interest, despite the presence of high-persistence and spurious critical points in  $\tilde{\rho}$ . In the following, we use  $G$  to extract repulsive interactions.

### 4.3 Repulsive Bond-Cycles

Following the characterization given in Sec. 3.1, a repulsive bond cycle is given as a cycle of minimal length in the Bonding Graph. Thus, we need to compute the shortest path between two bonded atoms by omitting their bond, as illustrated in Fig. 10. To do so, we weight the edges  $E$  of the graph  $G$  by the Euclidean distance of the incident nodes. Note that all the weights are positive. Let us consider a node representing a 1-saddle  $S \in N$  and its adjacent minima  $M_1, M_2 \in N$ . The two edges

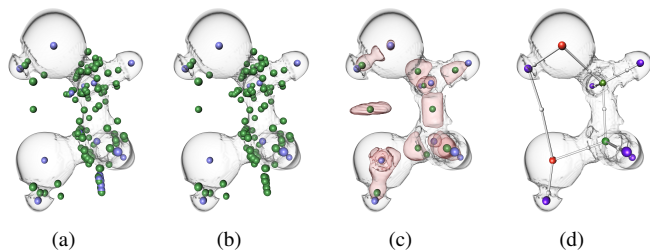


Fig. 7. Illustration of the different steps of our analysis algorithm: (a) low-persistence pairs are cancelled; (b) high-persistence minimum-saddle pairs are removed; (c) high-persistence saddle-saddle pairs are filtered out based on the interaction site extraction; (d) final bonding graph.

connecting these three nodes are denoted by  $\mathbf{E}_1, \mathbf{E}_2 \in E$ . We apply Dijkstra’s algorithm [17] to compute the shortest path  $P \subset E \setminus \{\mathbf{E}_1, \mathbf{E}_2\}$  between  $M_1$  and  $M_2$  in  $G$  omitting the edges  $\mathbf{E}_1$  and  $\mathbf{E}_2$ . If it exists, a molecular repulsive cycle was found. We mark all minima covered by the shortest path and collect all bonding saddles connecting the minima. We assign to each of these saddles an identifier indicating that all of them belong to the same repulsive interaction. Note that several saddles can describe the same cycle, i.e., all minima covered by two cycles coincide. In this case, we only keep one representative of this cycle.

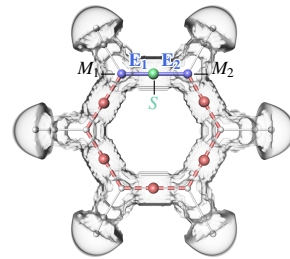


Fig. 10. Illustration of the repulsive bond-cycle extraction.

## 5 RESULTS AND DISCUSSION

In this section, we present experimental results of our analysis algorithm obtained with a C++ implementation on a computer with an i7 CPU (2.8 GHz) and 16Gb of RAM. We investigate a variety of molecular systems obtained through quantum chemistry simulation with the Gaussian program and represented by regular grid data. In all of experiments, our analysis took in general a few seconds to compute and at most 8 minutes for the DNA helix data-set ( $170 \times 178 \times 183$ ).

### 5.1 Validation

Fig. 8 shows the bonding graph computed by our approach on a variety of simple molecular systems. Our analysis indeed reveals the covalent bonds of these systems. Moreover, it also reveals the noncovalent interactions responsible for the bonding of several dimers. In particular, our analysis reveals that two hydrogen bonds (noncovalent



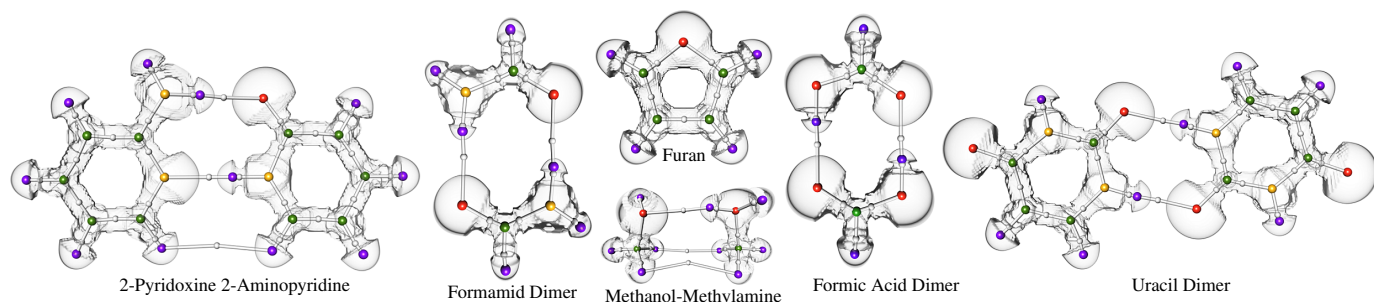


Fig. 8. Bonding Graphs of simple molecular systems. Oxygen, nitrogen, carbon, and hydrogen are shown as red, yellow, green, and purple spheres. Bonding saddles are depicted as white spheres. The Bonding Graph captures covalent and noncovalent bondings (white lines).

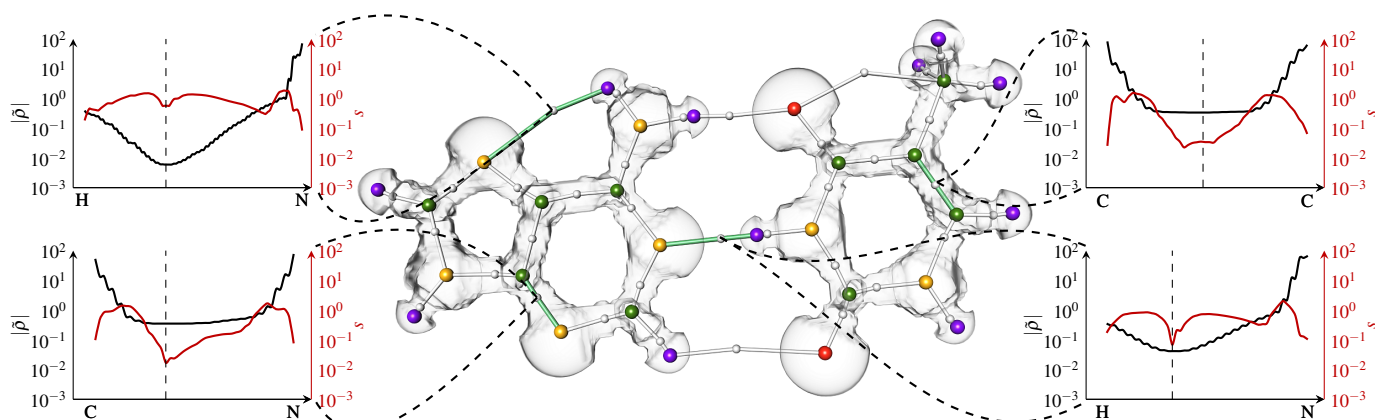


Fig. 9. Quantitative analysis of attractive interactions in the A-T DNA base pair. Thanks to the robust extraction of attractive bonds, each type of interaction can be investigated in the light of several quantum chemistry measures. By restricting the analysis of these measures to the separatrix representing each bond, our analysis enables to project these complex 3D informations down to easily readable 1D plots, revealing distinct characteristic behaviors for covalent bonds (bottom-left, upper-right), hydrogen bonds (bottom-right) or Van der Waals interactions (upper-left). The bond critical points are highlighted as dashed lines in the 1D plots.

bonds linking a hydrogen to a heavier atom) are involved in the bonding of the Pyridoxine and Aminopyridine (leftmost), while a van der Waals attractive force bonds the bottom two hydrogen atoms together. Each of these bonding graphs reveal a covalent and noncovalent bonding structure. The results are based on the selected kinks present in the respective persistence curves (Sections 4.1 and 4.2). Note that these kinks vary in strength as can be seen in the supplementing material.

## 5.2 Quantitative Analysis and Visual Exploration

Since our technique allows for a robust extraction of covalent and non-covalent interactions, it also enables to enumerate, classify and investigate such features from a quantitative point of view.

**Bond Investigation and Classification** In addition to the enumeration of the bonding interactions, our analysis enables new investigation capabilities on a per interaction basis. In particular, one can further analyze various quantum chemistry measures for each extracted

bond since our analysis provides a concrete geometrical representation for each interaction. Fig. 9 illustrates such an analysis where several quantities ( $|\bar{\rho}|$  and  $s$ ) were evaluated along the bond paths representing two covalent bonds (lower left and upper right) and noncovalent bonds (upper left and lower right). For each interaction, such a procedure enables to project these complex 3D information down to easily readable 1D plots revealing the evolution of these measures along the bond path. In particular, these plots reveal a characteristic behavior for covalent bonds. A high-valued flat plateau of electron population ( $|\bar{\rho}|$ ) is located in the middle of the bond. Moreover,  $s$  indicates a dominant minimum in the vicinity of the bond critical point. In contrast, noncovalent bonds reveal a more subtle electronic structure. In the vicinity of the bonding critical point, this analysis reveals that a minimum of electron population ( $|\bar{\rho}|$ ) is achieved at distinct values for hydrogen bonds ( $10^{-1}$ , bottom right) and van der Waals ( $10^{-2}$ , upper left) interactions; which is one to two orders of magnitude lower than covalent bonds. As proposed by Bader [1], attractive bonds can be classified according to the  $|\bar{\rho}|$ -value of their bonding critical point. We make use of this property to classify attractive bonds in the remainder.

**Steric Repulsion** Fig. 12 illustrates the extraction of repulsive bond cycles on the Pyridoxine-Aminopyridine. Steric repulsion is induced by a closed chain of atoms. As a first approach, we integrate  $|\bar{\rho}|$  along each cycle to assess its strength. This analysis reveals strong steric repulsions induced by the cycles formed by covalent bonds only (left and right orange cycles). In contrast, very weak steric repulsion is induced by the bottom cycle. This difference is caused by the presence of the electrons involved in covalent interactions and their absence in weak-energy van der Waals bond (bottom hydrogens). This insight is also confirmed by the  $|\bar{\rho}|$ -color-coding of the numerical streamline integration in  $\nabla\bar{\rho}$ .

**Complex Molecular Systems** Fig. 1 shows the extraction of attractive covalent and noncovalent bonds on a folded  $\beta$ -sheet polipep-

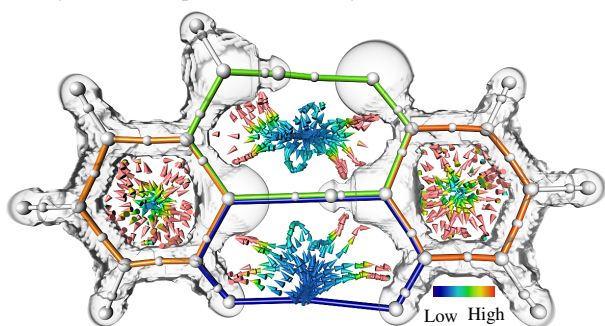


Fig. 12. Repulsive bond cycles of the Pyridoxine-Aminopyridine. Integrating  $|\bar{\rho}|$  along them enables to assess the strength of the steric repulsion, as confirmed visually by the  $|\bar{\rho}|$ -color-coding of  $\nabla\bar{\rho}$  (arrows).

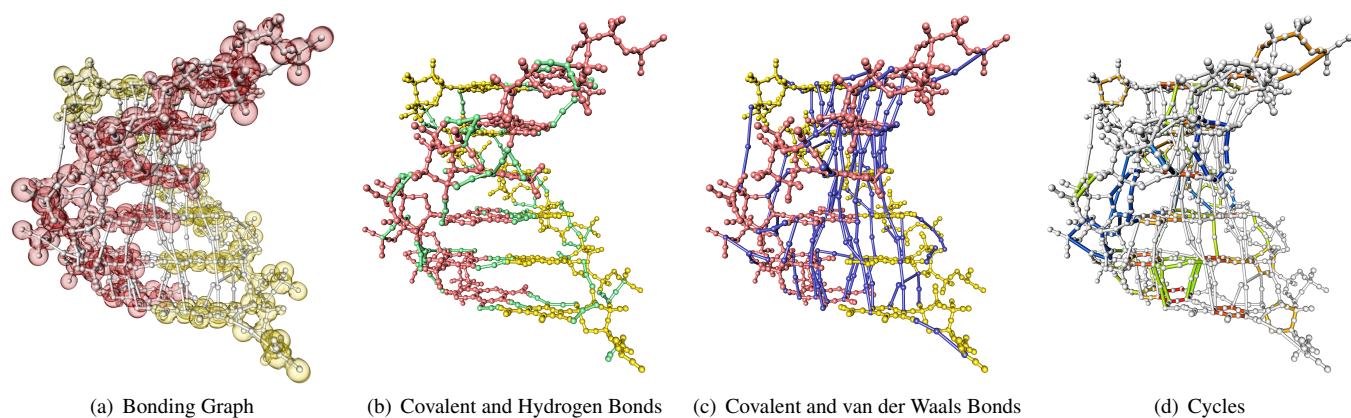


Fig. 11. Quantitative exploration of the nucleic acid double helix data-set. Our analysis enables to enumerate and classify noncovalent interactions (Fig. 13), yielding query-based visualizations (b, c) revealing the role of each type of interaction in the helicoidal structure of DNA. A multi-scale exploration of the repulsive bond cycles (d) reveals high repulsions in the center of the helix and weaker repulsion on its outer boundary.

tide. Our analysis enables to classify them according to the  $|\bar{\rho}|$ -value of their bond critical points. This classification highlights the noncovalent bonds responsible for the folding of this molecule. While hydrogen bonds (cyan) are involved in this folding, our analysis additionally reveals that van der Waals attractions (dark blue) also play a structural role to enforce the stability of this folded conformation. This distinction is confirmed by the continuous flow of  $\nabla\bar{\rho}$  (right inset), which exhibits the dispersed behavior of van der Waals attractions. Moreover, as highlighted with the black frame, our analysis reveals a repeating pattern of noncovalent bonds, which corresponds to the decomposition of the molecule in its elementary amino-acids.

Fig. 11 illustrates an exploration of the molecular interactions within a portion of a nucleic acid double helix found in DNA. Due to the complexity of this data-set, our algorithm extracts a large number of interactions between the two parts of the helix, highlighted in yellow and red (Fig. 11(a)). To reveal the structural role of these interactions at a global level, we perform a quantitative analysis of each of the extracted bonds. In particular, plotting the number of bonds as a function of  $|\bar{\rho}|$  (Fig. 13 left) indicates three distinct ascending modes. The first two modes (blue and green) are separated by a small-scale kink (see inset) while the last two modes (green and black) are separated by a large-scale kink. As suggested by our initial quantitative analysis on the A-T DNA base (Fig. 9), each portion of the curve (blue, green and black) correspond to a specific type of interactions: van der Waals, hydrogen bonds and covalent bonds, respectively. Thus, this curve enables to select thresholds for a query-based exploration of the noncovalent bonds. In Fig. 11(b), van der Waals attractions were removed while in Fig. 11(c) hydrogen-bonds were filtered out. Covalent bonds appear in red and yellow in these figures. These visualizations provide two complementary global insights on the helicoidal structure of this molecular system. In particular, Fig. 11(b) reveals that hydrogen bonds are mostly located in the planes orthogonal to the helix axis while van der Waals attractions follow the axis-direction. This confirms that the torsional stiffness of the double helix, which influences the circularisation of DNA, is mostly governed by hydrogen bonding, whereas the axial stiffness, which characterizes the wrapping properties of DNA, is mostly governed by weaker van der Waals attractions. A similar query-based visualization can be carried out regarding the steric repulsions. Since our analysis enables to quantify the strength of repulsive cycles (cf. previous paragraph), the evolution of their number can be plotted in function of their strength (Fig. 13 right). From this curve, the user can select relevant intervals (blue, green, red) and have a direct feedback in the visualization of the corresponding steric repulsions (Fig. 11(d)). This enables a multi-scale exploration of such features, which indicates a strong repulsion in the center of the base pairs (red cycles on the horizontal steps of the helix) and weaker repulsion on the outer boundary of the helix (blue cycles).

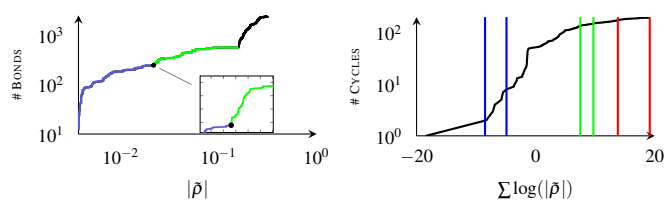


Fig. 13. Supporting plots for Fig. 11. Shown are the number of bonds and cycles present in the DNA example as a function of  $|\bar{\rho}|$ .

## 6 CONCLUSION

In this paper, we have presented a new approach for the visualization, enumeration, classification and investigation of covalent and noncovalent interactions in chemical systems. We showed how to isolate subtle chemical interactions (such as noncovalent bonds) despite the topological complexity of the input scalar fields. In contrast to existing approaches based on a qualitative visualization of the reduced gradient  $s$  [30], our combinatorial approach allows to derive quantitative analyses of these interactions. This enables further query-based visualizations revealing the underlying quantum chemistry governing complex molecular systems. With this tool, domain experts can now interactively investigate the mechanics driving molecular folding or bonding. Moreover, atoms and bonds can now be explicitly related to attractive and repulsive interactions – a tedious task which was manually carried out by domain experts before.

A current limitation of our approach regards repulsive bond cycles. While each of these can be robustly extracted with our approach, the interactions of these cycles together is not described by our visualizations. However, addressing this issue requires further theoretical investigations at the chemical level.

While our approach analyzes steady simulations, several quantum processes have also a time-dependent behavior. Tracking the creation and deletion of molecular interactions over time is therefore a stimulating research direction that finds a direct application in chemical design tasks such as pharmaceuticals. Our algorithm enables to isolate the chemical entities (atoms and bonds) responsible for molecular interactions. Consequently, we would like to extend our approach in the future by investigating the energetic implications of each interaction in processes such as molecular folding or bonding.

## ACKNOWLEDGMENTS

This research is supported and partially funded by the RTRA Digiteo unTopoVis project (2012-063D) and the framework of CALSIMLAB, which is supported by the public grant ANR-11-LABX-0037-01 overseen by the French National Research Agency (ANR) as part of the Investissements d’Avenir program (ANR-11-IDEX-0004-02).

## REFERENCES

- [1] R. Bader. *Atoms in Molecules: A Quantum Theory*. Oxford University Press, 1994.
- [2] C. Bajaj, V. Pascucci, A. Shamir, R. Holt, and A. Netravali. Dynamic maintenance and visualization of molecular surfaces. *Discrete Applied Mathematics*, 2003.
- [3] U. Bauer, C. Lange, and M. Wardetzky. Optimal topological simplification of discrete functions on surfaces. *Disc. & Comp. Geom.*, 2012.
- [4] D. Baum and H. Hege. A point-matching based algorithm for 3D surface alignment of drug-sized molecules. 2006.
- [5] K. Beketayev, G. Weber, M. Haranczyk, P. Bremer, M. Hlawitschka, and B. Hamann. Topology-based visualization of transformation pathways in complex chemical systems. *CGF (Proc. of EuroVis)*, 2011.
- [6] S. Biasotti, D. Giorgi, M. Spagnuolo, and B. Falcidieno. Reeb graphs for shape analysis and applications. *Theoretical Computer Science*, 2008.
- [7] H. Carr and D. Duke. Joint contour nets: Computation and properties. In *Proc. of PacificVis*, 2013.
- [8] H. Carr, J. Snoeyink, and U. Axen. Computing contour trees in all dimensions. In *Proc. of ACM SODA*, pages 918–926, 2000.
- [9] H. Carr, J. Snoeyink, and M. van de Panne. Simplifying flexible isosurfaces using local geometric measures. In *Proc. of IEEE VIS*, 2004.
- [10] F. Cazals, F. Chazal, and T. Lewiner. Molecular shape analysis based upon the Morse-Smale complex and the Connolly function. In *Proc. of ACM Symp. on Comp. Geom.*, 2003.
- [11] R. Chaudret, B. de Courcy, J. Contreras-Garcia, E. Gloaguen, A. Zehacker-Rentien, M. Mons, and J.-P. Piquemal. Unraveling non-covalent interactions within flexible biomolecules: from electron density topology to gas phase spectroscopy. *Physical Chemistry Chemical Physics*, pages –, 2014.
- [12] M. Chavent, A. Vanel, A. Tek, B. Levy, S. Robert, B. Raffin, and M. Baaden. GPU-accelerated atom and dynamic bond visualization using hyperballs: a unified algorithm for balls, sticks and hyperboloids. *Journal of Computational Chemistry*, 2011.
- [13] G. Cipriano and M. Gleicher. Molecular surface abstraction. *IEEE TVCG (Proc. of IEEE VIS)*, 2007.
- [14] A. J. Cohen, P. Mori-Sánchez, and W. Yang. Challenges for density functional theory. *Chemical Reviews*, 112(1):289–320, 2012.
- [15] M. Connolly. Analytical molecular surface calculation. *Journal of Applied Crystallography*, 1983.
- [16] J. Contreras-Garcia, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, and W. Yang. Nciplot: A program for plotting non-covalent interaction regions. *J. Chem. Theo. and Comp.*, 7(3), 2011.
- [17] E. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.
- [18] H. Doraiswamy and V. Natarajan. Computing Reeb graphs as a union of contour trees. *IEEE TVCG*, 2013.
- [19] H. Edelsbrunner and J. Harer. *Computational Topology. An Introduction*. Amer. Math. Society, Jan. 2010.
- [20] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Disc. & Comp. Geom.*, 28:511–533, 2002.
- [21] S. Fournais, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T. Oestergaard Soerensen. The electron density is smooth away from the nuclei. *Communications in Mathematical Physics*, 228(3), 2002.
- [22] J. L. Gázquez, J. del Campo, S. Trickey, R. J. Alvarez-Mendez, and A. Vela. Analysis of the generalized gradient approximation for the exchange energy. In *Concepts and Methods in Modern Theoretical Chemistry*, volume 1, pages 295–311. CRC Press, 2013.
- [23] N. Gillet, R. Chaudret, J. Contreras-Garcia, W. Yang, B. Silvi, and J. Piquemal. Coupling quantum interpretative techniques: Another look at chemical mechanisms in organic reactions. *Journal of Chemical Theory and Computation*, 2012.
- [24] J. Greer and B. Bush. Macromolecular shape and surface maps by solvent exclusion. *Proceedings of the National Academy of Science*, 1978.
- [25] D. Günther, J. Reininghaus, S. Prohaska, T. Weinkauff, and H. Hege. Efficient computation of a hierarchy of discrete 3D gradient vector fields. In *Proc. of TopInVis*, 2011.
- [26] D. Günther, J. Reininghaus, H. Wagner, and I. Hotz. Efficient computation of 3D Morse-Smale complexes and persistent homology using discrete Morse theory. *The Visual Computer*, 28:959–969, 2012.
- [27] A. Gyulassy, P.-T. Bremer, B. Hamann, and P. Pascucci. A practical approach to Morse-Smale complex computation: scalability and generality. *IEEE TVCG (Proc. of IEEE VIS)*, pages 1619–1626, 2008.
- [28] A. Gyulassy, V. Natarajan, V. Pascucci, P. Bremer, and B. Hamann. A topological approach to simplification of three-dimensional scalar functions. *IEEE TVCG*, 2006.
- [29] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *The American Physical Society*, 1964.
- [30] E. Johnson, S. Keinan, P. Mori-Sanchez, A. Contreras-Garcia, J. Cohen, and W. Yang. Revealing noncovalent interactions. *Journal of the American Chemical Society*, 2010.
- [31] M. Joswig and M. E. Pfetsch. Computing optimal Morse matchings. *SIAM Journal on Discrete Mathematics*, 20(1):11–25, 2006.
- [32] M. Krone, K. Bidmon, and T. Ertl. Interactive visualization of molecular surface dynamics. *IEEE TVCG (Proc. of IEEE VIS)*, 2009.
- [33] M. Krone, M. Falk, S. Rehm, J. Pleiss, and T. Ertl. Interactive exploration of protein cavities. *CGF (Proc. of EuroVis)*, 2011.
- [34] M. Krone, G. Reina, C. Schulz, T. Kulschewski, J. Pleiss, and T. Ertl. Interactive extraction and tracking of biomolecular surface features. *CGF (Proc. of EuroVis)*, 2013.
- [35] J. R. Lane, J. Contreras-Garcia, J.-P. Piquemal, B. J. Miller, and H. G. Kjaergaard. Are bond critical points really critical for hydrogen bonding? *Journal of Chemical Theory and Computation*, 9(8):3263–3266, 2013.
- [36] G. Lewis. The Atom and the Molecule. *Am. Chem. Soc.*, 1916.
- [37] N. Lindow, D. Baum, and H. Hege. Voronoi-based extraction and visualization of molecular paths. *IEEE TVCG (Proc. of IEEE VIS)*, 2011.
- [38] C. Matta and R. Boyd. *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*. Wiley, 2007.
- [39] J. Milnor. *Morse Theory*. Princeton University Press, 1963.
- [40] J. Milnor. *Lectures on the H-Cobordism Theorem*. Princeton University Press, 1965.
- [41] M. Morse. *The Calculus of Variations in the Large*. Number v. 18 in Colloquium Publications - American Mathematical Society. American Mathematical Society, 1934.
- [42] V. Natarajan, Y. Wang, P. Bremer, V. Pascucci, and B. Hamann. Segmenting molecular surfaces. *Computer Aided Geometric Design*, 2006.
- [43] S. O'Donoghue, D. Goodsell, A. Frangakis, F. Jossinet, R. Laskowski, M. Nilges, H. Saibil, A. Schafferhans, R. Wade, W. E., and O. A. Visualization of macromolecular structures. *Nature Methods*, 2010.
- [44] S. Parsa. A deterministic  $O(m \log(m))$  time algorithm for the Reeb graph. In *Proc. of ACM Symp. on Comp. Geom.*, 2012.
- [45] J. Parulek and A. Brambilla. Fast blending scheme for molecular surface representation. *IEEE TVCG (Proc. of IEEE VIS)*, 2013.
- [46] J. Parulek, C. Turkey, N. Reuter, and I. Viola. Implicit surfaces for interactive graph based cavity analysis of molecular simulations. In *Proc. of BioVis*, 2012.
- [47] L. Pauling. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*. Cornell University Press, 1960.
- [48] J. Reininghaus, D. Günther, I. Hotz, T. Weinkauff, and H.-P. Seidel. Combinatorial gradient fields for 2D images with empirically convergent separatrices, 2012. Arxiv:1208.6523v1.
- [49] F. Richards. Areas, volumes, packing, and protein structure. *Annual Review of Biophysics and Bioengineering*, 1977.
- [50] V. Robins, P. J. Wood, and A. P. Sheppard. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Trans. on PAMI*, 33(8):1646–1658, 2011.
- [51] S. Smale. On gradient dynamical systems. *The Annals of Mathematics*, 74:199–206, 1961.
- [52] H. Theisel, T. Weinkauff, H.-C. Hege, and H.-P. Seidel. Saddle connectors - an approach to visualizing the topological skeleton of complex 3d vector fields. *IEEE TVCG (Proc. of IEEE VIS)*, pages 225–232, October 2003.
- [53] J. Tierny, A. Gyulassy, E. Simon, and V. Pascucci. Loop surgery for volumetric meshes: Reeb graphs reduced to contour trees. *IEEE TVCG (Proc. of IEEE VIS)*, 15:1177–1184, 2009.
- [54] M. van der Zwan, W. Lueks, H. Bekker, and T. Isenberg. Illustrative molecular visualization with continuous abstraction. *CGF (Proc. of EuroVis)*, 2011.
- [55] G. Weber, S. Dillard, H. Carr, V. Pascucci, and B. Hamann. Topology-controlled volume rendering. *IEEE TVCG*, 2007.
- [56] T. Weinkauff. *Extraction of Topological Structures in 2D and 3D Vector Fields*. PhD thesis, University Magdeburg, 2008.
- [57] K. Yosida. *Functional Analysis*. Classics in Mathematics. Cambridge University Press, 1995.
- [58] X. Zhang and C. Bajaj. Extraction, quantification and visualization of protein pockets. In *Proc. of IEEE CSBC*, 2007.