Voronota-LT: efficient, flexible and solvent-aware tessellation-based analysis of atomic interactions

Kliment Olechnovič, Sergei Grudinin

Voronota-LT: efficient, flexible and solvent-aware tessellation-based analysis of atomic interactions

Kliment Olechnovič[1,2,*] and Sergei Grudinin[1]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France and [2]Institute of Biotechnology, Life Sciences Center, Vilnius University, 10257 Vilnius, Lithuania

[*]Corresponding author. kliment.olechnovic@univ-grenoble-alpes.fr, kliment.olechnovic@bti.vu.lt

Abstract

Motivation: In the fields of structural biology and bioinformatics, understanding molecular interactions is paramount. However, existing advanced geometric methods for describing interatomic contacts considering full structural context have typically demanded substantial computational resources, hindering their practical application. Given the ever-growing volume of structural data, there is an urgent need for more efficient tools for interaction analysis.

Results: We present Voronota-LT, a new efficient method tailored for computing Voronoi tessellation-based atom-atom contacts within the solvent-accessible surface of molecular structures. Voronota-LT delivers results that correlate highly with the original Voronota method, but does it significantly faster. The new method is parallelizable and capable of selectively targeting specific interface areas within molecular complexes. While offering high execution speed, Voronota-LT provides a comprehensive description of every interatomic interaction, taking full account of the relevant structural context.

Availability and Implementation: Voronota-LT software is freely available at https://kliment-olechnovic.github.io/voronota/expansion_lt/.

Key words: structural bioinformatics, radical Voronoi tessellation, Laguerre-Voronoi diagram, solvent-accessible surface, atom-atom contact areas

1. Introduction

There are multiple ways to computationally define and analyze interatomic contacts in a three-dimensional (3D) model of a molecular structure, but most studies in structural bioinformatics only rely on calculating and interpreting atom-atom distances. This is not surprising, as such analysis is usually fast and easy to implement. However, there are alternative approaches that are more descriptive. For example, the Voronoi tessellation (Voronoi, 1908) and its weighted variants that can consider atomic radii, Laguerre-Voronoi diagram (Imai et al., 1985; Aurenhammer, 1987) and additively weighted Voronoi tessellation (Goede et al., 1997; Kim et al., 2005), were known for a long time and used to investigate various properties of molecular structures (Poupon, 2004; Cazals, 2006).

While distance calculation for a pair of atoms does not depend on the other surrounding atoms, the Voronoi tessellation-based interaction analysis accounts for the structural neighbors that may affect a given contact (Esque et al., 2011). Implicit interactions with solvent can be added by constraining Voronoi cells inside the solvent-accessible surface of a molecule (Goede et al., 1997). The faces of the constrained Voronoi cells can be described quantitatively by calculating their areas after on-sphere projection (McConkey et al., 2002) or directly (Olechnovič and Venclovas, 2017). Tessellation-derived contact areas can then be used to define machine learning-based scoring functions that estimate, for example, pseudo-energy for every interaction (McConkey et al., 2003; Verdonk et al., 2016; Olechnovič and Venclovas, 2017, 2023; Igashov et al., 2021a,b). Because such contact area-based scores correspond to the partitioning of space between atoms, they can be added up to obtain to a global score value that accounts for the packing arrangement of atoms. In contrast, pairwise distance-based scoring functions produce values that are inherently non-additive from the perspective of statistical physics (Ben-Naim, 1997).

The most direct approach to constraining Voronoi tessellation-derived atom-atom contacts inside the solvent-accessible surface is simply cutting Voronoi cell faces by the solvent-accessible surface. Surprisingly, to the best of our knowledge, only one publicly available software tool uses this direct approach for the analysis of molecular structures — Voronota (Olechnovič and Venclovas, 2014, 2021). There are recent testimonies to the usefulness of Voronota-computed contacts: they were used as the basis for the best-performing method (Olechnovič et al., 2023) in the scoring challenge in the recent CASP15-CAPRI experiment (Lensink et al., 2023); they comprised the inter-chain interface feature descriptors that were assessed to be some of the most useful for discriminating physiological from non-physiological interfaces in structures of protein complexes (Schweke et al., 2023). It was also reported that Voronota computes the global Voronoi tessellation of balls relatively very rapidly when the input is biomolecular structural data (Lee et al., 2022). Nevertheless, the total time

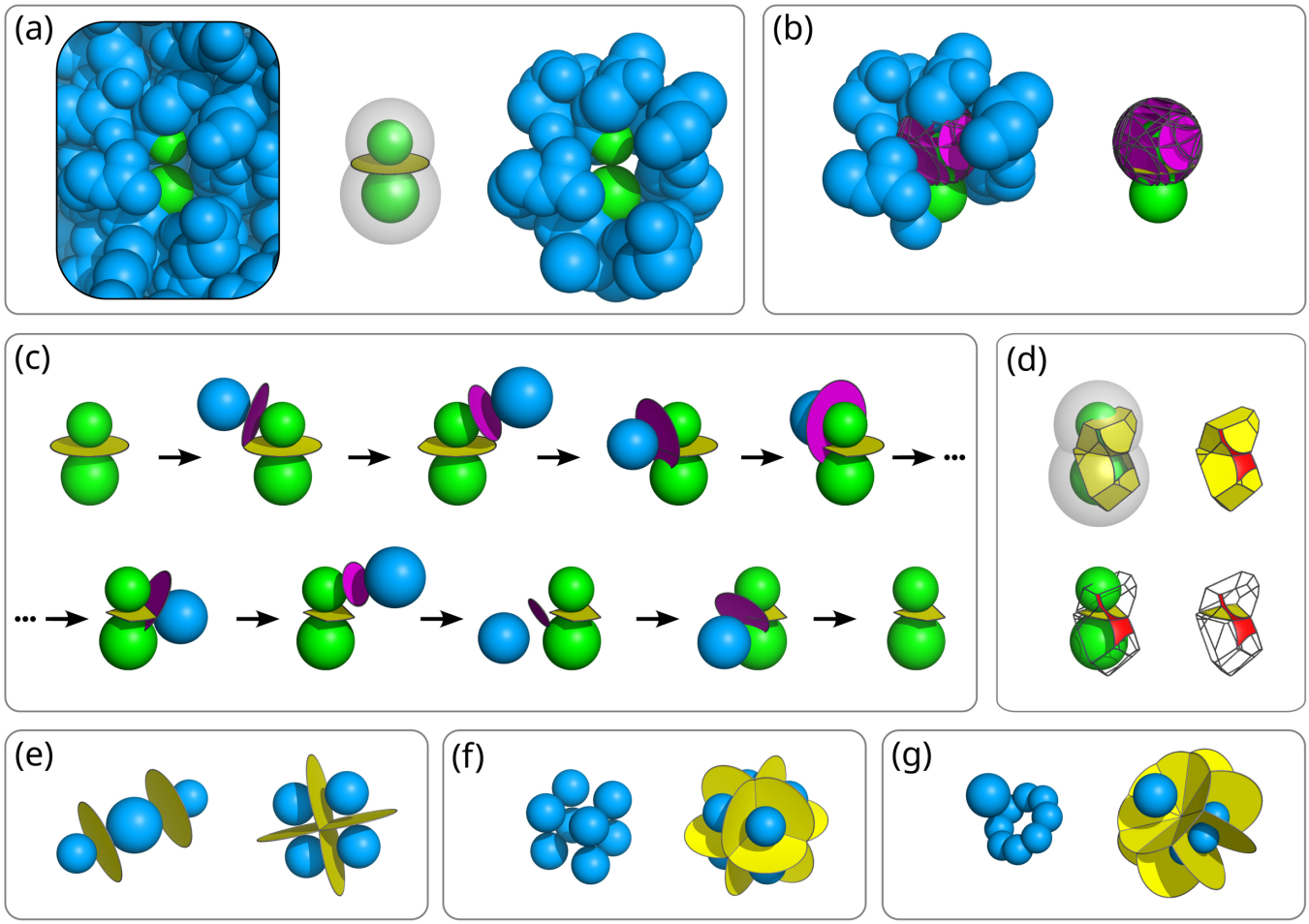

Fig. 1: Algorithmic aspects of Voronota-LT. **(a)** A pair of possibly contacting atoms (colored green) surrounded by other atoms (colored blue) in a protein structure (left); the maximum-area contact defined as a disk produced by intersecting two probe-expanded spheres (middle); neighboring blue atoms that are close enough to have contact disks with either of the green atoms (right). **(b)** All close-enough neighbors of the first green atom (left), and the corresponding contact disks (right). **(c)** Illustration of the iterative contact cutting process, starting with the maximum-area contact defined as a disk and ending with the final contact that cannot be cut anymore. **(d)** Illustrations of the constrained Laguerre-Voronoi cells formed by all the produced contacts of the green atoms. The exposed circular boundaries of the contacts define solvent-accessible surface patches (colored red) on the probe-expanded spheres. **(e-g)** Examples of contacts constructed correctly for arrangements of balls that induce degenerate Laguerre-Voronoi cells: balls with centers on a line and on a square lattice (e); balls on a cubic lattice (f), balls in a rotationally symmetrical arrangement (g).

that Voronota needs to compute contacts can still reach several seconds for a moderately-sized biomolecular structure of about 5,000 atoms.

In this work, we present a drastically more time-efficient approach to constructing Voronoi tessellation-derived atom-atom contacts for macromolecules with conservatively defined solvent-accessibility surfaces. The new method is called "Voronota-LT" (pronounced as "Voronota lite") and is freely available at https://kliment-olechnovic.github.io/voronota/expansion_lt/.

2. Methods

2.1. Definition of an atom-atom contact

Given an input molecular structure, we interpret atoms as balls of van der Waals radii. For every atom, we define a *contact sphere* around it by augmenting the atomic ball radius with the solvent probe radius (commonly set to 1.4 Å). We say that two atoms are *close enough* to have a physically relevant contact if their contact spheres intersect, that is, if a solvent probe cannot fit between the two atomic balls. When considering such two atoms outside of any neighboring structural context, we define the contact between the atoms to be the *disk* defined by the intersection of the two contact spheres (Fig. 1a). When atoms have multiple neighbors, there can be multiple contact disks that can intersect (Fig. 1b).

Every contact disk lies on a *radical plane* defined by an intersection of two spheres, and intersections of radical planes are known to define the *radical Voronoi tessellation* (also known as the *Laguerre-Voronoi tessellation* (Imai et al., 1985) or the *power diagram* (Aurenhammer, 1987)). Therefore, the intersections of a contact disk with other contact disks can produce a portion of a radical Voronoi cell face that is constrained inside the initial contact disk. Such a constrained contact surface respects the

geometric partitioning of space between the atomic balls. It also implicitly considers the presence of solvent by not extending outside of the solvent-accessible surface of a molecular structure. Thus, we assume it is a valid representation of an interatomic interaction.

2.2. Atom-atom contact construction algorithm

Given a pair of atoms a and b that have a contact disk, without loss of generality, let us assume that a has a smaller set of close-enough neighbors (for example, the upper green ball in Fig. 1a and 1b). This set contains all the possible balls that are close-enough Voronoi neighbors to both atoms. We set the initial contact representation to be the *full contact disk* between a and b. We then iteratively cut it with every contact disk defined for a and its close-enough neighbors. Figure 1c illustrates this process. After every cut, we are left with the portion of the contact surface that is in the same half-space of the cutting disk plane as a. We stop the process when there is no surface left (meaning that a and b are not in contact anymore), or after all the neighbors of a were processed.

While our algorithm naively follows the definition of the radical Voronoi tessellation without employing any clever heuristics, there are several aspects that make it practical. Firstly, in molecular structures, the density of the atomic packing tends to be not too far from the level allowed by the laws of physics. Therefore, the number of close-enough neighbors for every atom are bound by a constant value when using a solvent probe radius close to 1.4 Å. Secondly, we employ only computationally cheap geometric operations when performing intersection filtering and cutting operations. In the iterative portion of the algorithm, we initially represent the contact disk by its minimal surrounding hexagonal polygon, we cut that polygon iteratively, and only when the cutting is over we replace polygon line segments with circular arcs where needed. Thirdly, different contacts are constructed independently from each other, and the the final result does not depend on the order of the construction. Therefore, the algorithm is trivially parallelizable.

2.3. Constrained Voronoi cell characterization algorithm

We define the *constrained Voronoi cell* of an atom as the space region bounded by the contact surfaces and the contact sphere of that atom (Fig. 1d). Such a cell may be not fully closed because some contacts can have circular arc edges. Such arcs lie on the contact sphere surface and emborder regions of the solvent-accessible surface of the atom (colored in red in Fig. 1d), that is, parts of the contact sphere where a solvent probe center can be placed without being inside any neighboring Voronoi cell.

To compute the total solvent-accessible area corresponding to the constrained Voronoi cell, we use the fact that the area of a region on a sphere surface can be directly calculated from the solid angle subtended by the region from the sphere center (Todhunter, 1863). Instead of computing the solid angle of the solvent-accessible surface directly, we calculate the solid angles for all the atom-atom contact surfaces, sum them up, and subtract the sum from the total solid angle of a sphere that equals 4π steradians. As every our contact is a convex polygon with sides that are either line segments or circular arcs, we can compute every contact-subtended solid angle analytically. There is a pitfall in the radical Voronoi tessellation when the sphere's center lies outside the sphere's Voronoi cell, which can make the solid angles subtended by the cell faces from the sphere's center overlap. We handle such overlaps by inverting solid angle values (subtracting them from 4π) where needed.

Knowing the total solid angle of the solvent-accessible regions also allows us to compute the volume of the corresponding spherical sector inside the contact sphere. To get the total volume of the constrained Voronoi cell, we add the total sum of volumes of the pyramids formed by using the sphere's center as the apex and the contact surfaces as bases. In cases when the sphere's center lies outside the Voronoi cell, the pyramids can overlap. Therefore, we invert pyramid volumes (multiply them by -1) where needed.

We characterize Voronoi cells using only contact polygons without needing explicitly pre-computed Voronoi vertices or edges. It makes our approach robust in situations where input contains arrangements of balls that induce degenerate Voronoi vertices or edges, that is, vertices shared by more than four cells or edges shared by more than three cells (Fig. 1efg). Such situations can frequently occur in molecular structures as they often contain symmetrical parts.

2.4. Implementation

We implemented the contact construction and cell characterization algorithms in C++ as a header-only library and as a standalone command-line tool, both called "Voronota-LT". The parallelization is implemented using OpenMP (Dagum and Menon, 1998) and can be enabled during compilation. Apart from the optional usage of OpenMP, both the library and the command-line tool do not depend on anything but the C++ standard library.

Notably, Voronota-LT is completely independent from any code from the original Voronota software, called "vanilla Voronota" in this paper. However, we distribute Voronota-LT as an expansion part of the Voronota software package for two reasons: to enable other Voronota expansions to easily use the Voronota-LT library; to bundle the Voronota-LT command-line tool with the vanilla Voronota command-line tools that can produce input for Voronota-LT by parsing molecular structure files and assigning van der Waals radii to atoms.

In its default running regime, Voronota-LT computes all possible atom-atom contact areas, solvent-accessible areas and volumes. However, given an input list of atomic balls with labels that include chain and/or residue identifiers, there are options to limit Voronota-LT to compute only inter-residue or inter-chain contacts. The relevant contact areas computed in these limiting regimes are the same as when running in the default regime, but the total running time can be significantly lower because all non-relevant (intra-residue or intra-chain) contacts are not computed. Because atom-atom contact surfaces do not overlap, atom-atom contact areas can be directly summed into residue-residue or chain-chain contact areas according to the provided labeling of the input atomic balls.

2.5. Comparison of algorithms in Voronota-LT and in vanilla Voronota

The vanilla Voronota software (Olechnovič and Venclovas, 2014, 2021) employs a substantially different approach to constructing atom-atom contacts. Given an input set of balls, Voronota first computes all the vertices of the additively weighted Voronoi tessellation, and then uses the vertices to identify neighbors that are both close enough and have adjacent Voronoi cells. Then, every contact is constructed starting from the initial triangulated representation that corresponds to the intersection of

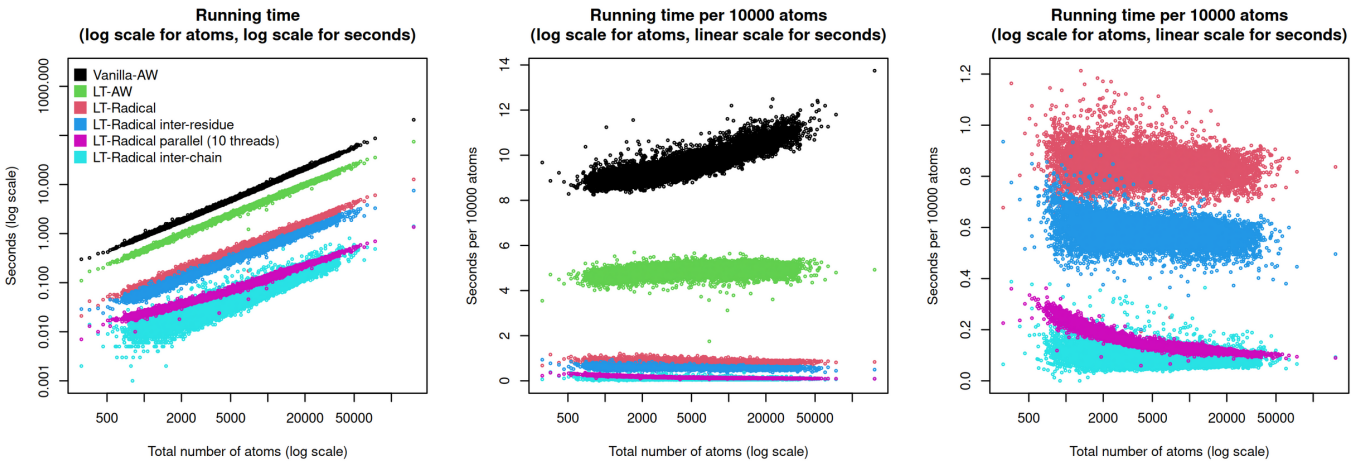

Fig. 2: Time efficiency of constructing contacts using vanilla Voronota (Vanilla-AW) and various modes of Voronota-LT (LT-*). Total running times (including input, computation and output routines) were used for every plot.

a contact sphere with a hyperboloidal surface (which is a plane when the two contacting atoms have the same radius). That triangulated representation is then iteratively updated by cutting it with neighboring hyperboloidal surfaces. The resulting contact surface can be non-convex and can have disconnected parts. The solvent-accessible surface is constructed from the triangulated representation of a contact sphere cut by neighboring contact spheres.

In comparison, Voronota-LT does not pre-compute Voronoi vertices, uses a much simpler representation of a contact surface, and computes all the areas and volumes analytically. Also, unlike vanilla Voronota, Voronota-LT is easily parallelizable.

The general Voronota-LT methodology for constructing contacts allows the usage of different forms of the Voronoi diagram. Therefore, we could additionally implement the construction of constrained triangulated contact surfaces that correspond to the additively-weighted Voronoi tessellation, like in vanilla Voronota. We called this variation "Voronota-LT-AW" ("AW" stands for "Additively Weighted"), while the default, a simpler and faster variation that uses the radical Voronoi tessellation is called "Voronota-LT-Radical" or just "Voronota-LT".

3. Benchmarking

3.1. Data

We benchmarked the Voronota-LT command-line tool on a diverse set of structures of multi-chain biological assemblies downloaded from the Protein Data Bank (PDB) (wwPDB consortium, 2019). We queried the PDB using RCSB.org GraphQL-based API (Rose et al., 2021) for protein and nucleic acid entities that belonged to PDB entries deposited before 2024-01-03 and where the first biological assembly contained two or more polymer chain instances (with up to 5000 polymeric residues overall) and at least one non-polymer entity instance. We took the PDB IDs of the entities that were cluster representatives at 30% sequences identity. For every PDB ID, we downloaded the first biological assembly in the mmCIF format and converted it to a labeled list of atomic balls with the Voronota software, taking only heavy (non-hydrogen) atoms. The van der Waals radii were assigned using the previously published atom type-to-radius mapping (Li and Nussinov, 1998; Olechnovič and Venclovas, 2021). We then filtered out structures with extreme clashes where two or more balls had exactly the same center. In total, we obtained 14,861 molecular structures with atom counts ranging from 256 to 151,585.

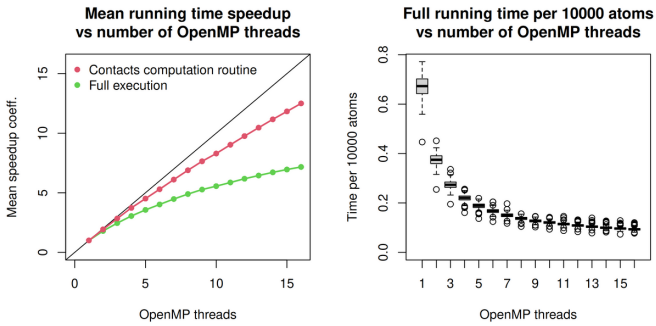

Fig. 3: Parallel speedup of Voronota-LT-Radical computations (left) and the full running time distributions (right) for the number of OpenMP threads ranging from 1 to 16.

3.2. Running time analysis

For every prepared input structure, we computed atom-atom contact areas with the Voronota-LT method in five different modes: "LT-Radical" (default mode, computing all atom-atom contacts areas and also computing all solvent-accessible surface areas), "LT-Radical parallel" using 10 OpenMP threads (computing the same areas as in the default mode), "LT-Radical inter-residue" (computing only contact areas between atoms from different residues), "LT-Radical inter-chain" (computing only contact areas between atoms from different chains), "LT-AW" (computing atom-atom contact areas that correspond to the additively-weighted Voronoi tessellation). We also ran vanilla Voronota software in its default mode (denoted as "Vanilla-AW") that computed atom-atom contacts and solvent-accessible surface areas. All the software was compiled and run on the same machine (20-core Intel® Xeon® Silver 4210R CPU @ 2.40GHz, Ubuntu Linux 23.04, GCC g++ compiler version 12.3.0).

We recorded the wall running time values and plotted them in Fig. 2. On average, when compared to vanilla Voronota, "LT-AW" is only two times faster, but "LT-Radical" is 12 times faster. The Voronota-LT-Radical modes that only computes inter-chain and inter-residue contacts are, respectively, 17.5 and 119 times faster than "Vanilla-AW". Moreover, as evident from the middle and the right parts of Fig. 2 that show running times per 10,000 input atoms, the efficiency of Voronota-LT (in any mode) does not decrease when the input size increases.

3.3. Parallelization efficiency analysis

The Voronota-LT-Radical executed in parallel using 10 threads performed, on average, 76 times faster than vanilla Voronota, but only 6.3 times faster than the unparallelized Voronota-LT-Radical (Fig. 2). This is not surprising because only the contacts computation procedure is effectively parallelized in Voronota-LT.

To further investigate the parallelization efficiency, we took the 500 largest input structures (with atom counts ranging from 22,169 to 151,585), processed them with Voronota-LT-Radical using different numbers of OpenMP threads, recorded wall running times, and calculated the average parallel speedup coefficients for every number of threads. We did it for both the total running times and the contacts computation procedure running times.

The speedup plot in Fig. 3 shows that the computation of the contacts is parallelized relatively efficiently, especially for a lower number of threads where the speedup is close to ideal. However, unsurprisingly, when increasing the number of threads, the total execution speedup grows much slower than the speedup of the contacts computation. The right plot in Fig. 3 shows what improvements of time per 10,000 atoms to expect from the parallelized Voronota-LT using different numbers of threads.

3.4. Correllations of areas computed by Voronota-LT and vanilla Voronota

The differences between van der Waals radii of different atoms in biological macromolecular structures are not extreme. Therefore, contact areas based on the radical and additively weighted Voronoi tessellations are expected to correlate highly. Thus, the contact areas computed by Voronota-LT should correlate highly with those from the vanilla Voronota. Moreover, when the van der Waals radii are set to the same constant value, the results of the Voronota-LT and the vanilla Voronota software tools should correlate ideally (there may be very slight differences attributed to the usage of floating point arithmetics in both tools and to the fact that vanilla Voronota always uses triangulated approximations of surfaces for computing areas). Another outcome may indicate there is something wrong with the Voronota-LT implementation.

In order to experimentally validate the results of Voronota-LT, we computed the Pearson correlation coefficients between differently produced sets of areas for every input structure. We summarized every distribution of coefficients by calculating the minimal ($\rho_{\min}$) and the average (ρ_{avg}) values. For a more convenient display, we rounded up those values to four significant digits.

When all the input ball radii were set to 1.8 Å, the vanilla Voronota and Voronota-LT contact areas correlated ideally ($\rho_{\min} = 1$, $\rho_{\text{avg}} = 1$). When the radii were not changed to be the same, the vanilla Voronota and Voronota-LT-AW contact areas correlated almost ideally ($\rho_{\min} = 0.999$, $\rho_{\text{avg}} = 1$). Vanilla Voronota and Voronota-LT-Radical contact areas correlated with $\rho_{\min} = 0.9609$ and $\rho_{\text{avg}} = 0.976$. Interestingly, the contact areas summarized on the residue-residue level correlated even more highly ($\rho_{\min} = 0.9947$, $\rho_{\text{avg}} = 0.999$). The summarized chain-chain areas correlated even better, the Pearson correlation coefficient calculated for all 67,700 interfaces in the benchmark dataset was greater than 0.9999.

Solvent-accessible surface areas computed by vanilla Voronota and Voronota-LT also correlated almost ideally ($\rho_{\min} = 0.9997$, $\rho_{\text{avg}} = 1$).

4. Application examples

4.1. Studying contact area variability in a structural ensemble

As we have established in the benchmarking study, Voronota-LT can compute detailed tessellation-derived atom-atom contact descriptors very rapidly. Thus, it can be used to efficiently process ensembles of multiple structural states of the same molecule, giving the ability to collect statistics of areas for every observed contact. The area of a contact depends not only on the distance between the corresponding interacting atoms but also on how the contact is influenced by the surrounding atoms. Therefore, calculating simple statistics of distributions of contact areas, for example, variance values, can help to characterize the variability of every interaction without disregarding the effects of its structural environment. Contact areas can also be used as weights when summarizing the variability of multiple interactions when characterizing interfaces inside, for example, a molecular complex.

Figure 4 presents an example of a simple contact area variability analysis. We took an NMR ensemble (PDB ID 1CIR) of 20 structural states from the study on chymotrypsin inhibitor 2 folding (Neira et al., 1996). It is a dimeric protein molecule, and we focused on the contacts in the inter-chain interface. We computed the contact areas and, for every unique contact, we calculated the mean and the variance values of the corresponding observed area distribution. We considered both atom-atom contact areas and the residue-residue level areas derived by summing atom-atom areas according to residue identifiers. We visualized the results by displaying inter-chain contact surfaces and coloring them by the corresponding variance-to-mean ratios (relative variances) that indicate how much every contact area is dispersed relative to its mean. We observed what contact areas varied the least and the most. In particular, the contact areas further from the solvent, in the interface core, were more stable. The most variable interface region involved interactions with a C-terminal tail of chain A.

Due to the speed of Voronota-LT, this type of analysis is highly scalable. It can be performed for much larger molecules with much more structural states, and it can be focused not only on inter-chain contacts but also on all contacts or any subset of contacts. Another benefit of the tessellation-based variability analysis is that structural states do not need to be superposed, that is, structurally aligned.

4.2. Studying contact changes in large structures from time-resolved cryo-EM experiments

Modern time-resolved cryo-EM experiments can provide structural models of big molecular complexes at various stages of biological processes. Analysis of such data can be challenging due to its scale and complexity. We present an example of how a tessellation-based analysis can help identify and describe changes in interactions between chains in a 70S *E. coli* ribosome that undergoes splitting

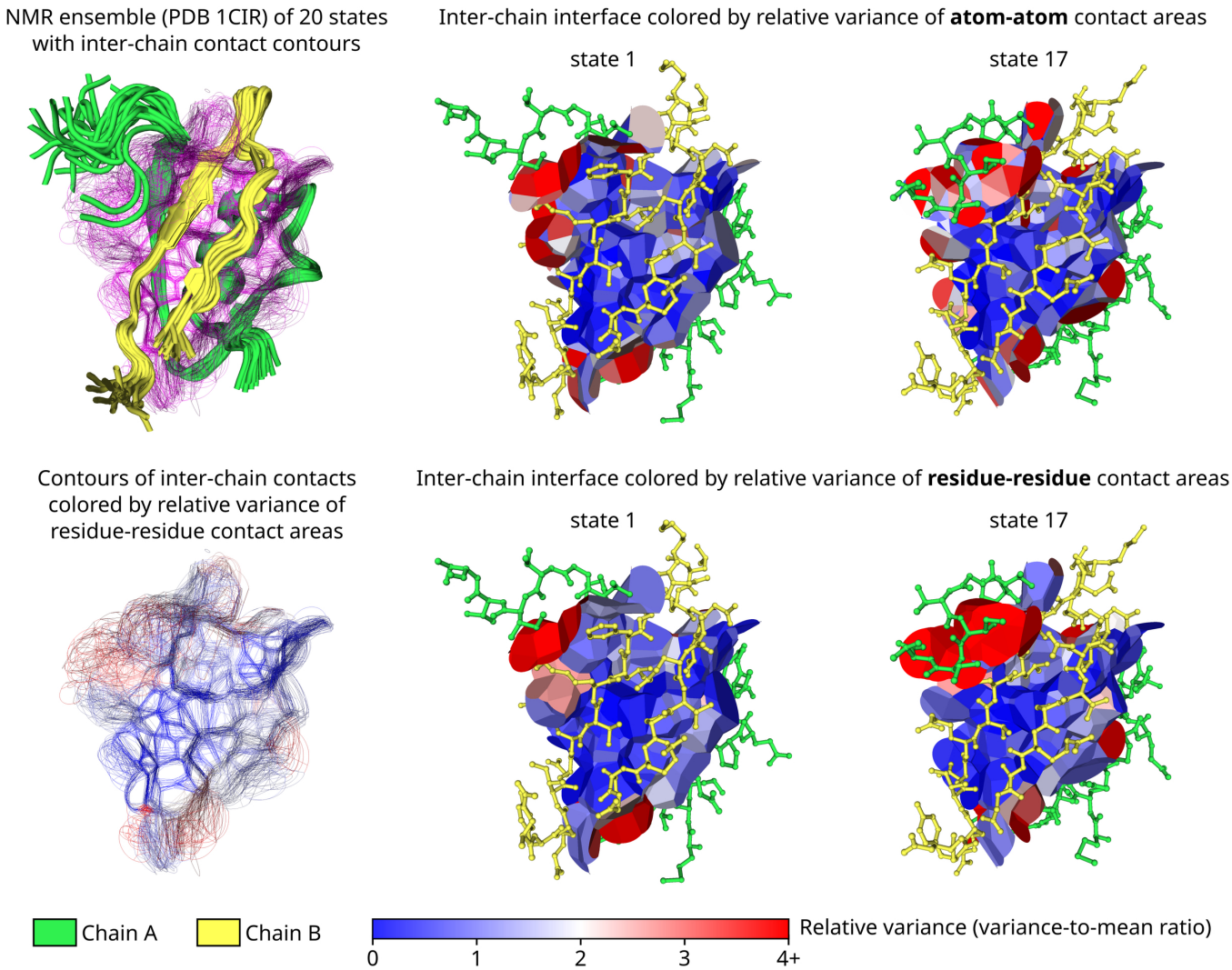

Fig. 4: Example application of Voronota-LT for studying inter-chain interface contact variability in an ensemble of dimeric protein structures. The analysed NMR ensemble was taken from the Protein Data Bank entry 1CIR deposited by the authors of the study on chymotrypsin inhibitor 2 folding (Neira et al., 1996). 3D visualizations were generated using Voronota-GL.

of its 50S and 30S subunits mediated by an external protein factor. We took structural data from the time-resolved cryo-EM study of HflX-mediated ribosome recycling (Bhattacharjee et al., 2023). We analysed differences between inter-chain interface contacts computed for two ribosome structures determined on an earlier (PDB ID 8G34) and on a later (PDB ID 8G38) stages of splitting (Fig. 5a). We computed inter-chain contact areas on the residue-residue level and calculated contact area differences between different structural states. Inter-chain interfaces inside the 50S and 30S subunits did not differ systematically, but the contact changes in the interfaces between 50S and 30S were dramatic (Fig. 5b). The tessellation-based analysis allowed us to describe the almost complete rearrangement of the contacts between the 50S and 30S subunits (Fig. 5c) , end the emergence of an additional interface involving the HflX factor in the later stage of splitting (Fig. 5d).

In the previous application examples, we used contact areas on the atom-atom or residue-residue levels. Voronota-LT can also summarize contact areas on the chain-chain level, which is particularly useful when exploring molecular complexes with many chains. For example, we computed inter-chain interface contact areas for the contracted (PDB ID 8C38) and the extended (PDB ID 8CPY) structural states of the cowpea chlorotic mottle virus capsid determined during the time-resolved cryo-EM study of viral dynamics (Harder et al., 2023). Each structure had 180 protein chains of the same sequence, the total number of atoms was 215,580 for 8C38 and 198,420 for 8CPY. Using 16 OpenMP threads, it took Voronota-LT about 0.6 and 0.5 seconds to compute all chain-chain interface areas for 8C38 and 8CPY, respectively. We used the computed results to evaluate the difference in chain packing (Fig. 6). The contracted variant had 786 inter-chain interfaces with a total area of 365,093 $Å^2$, while the extended variant had 390 inter-chain interfaces with a total area of 184,506 $Å^2$. Interestingly, histograms of the chain-chain interface areas indicate that some of the largest interface areas that occurred in

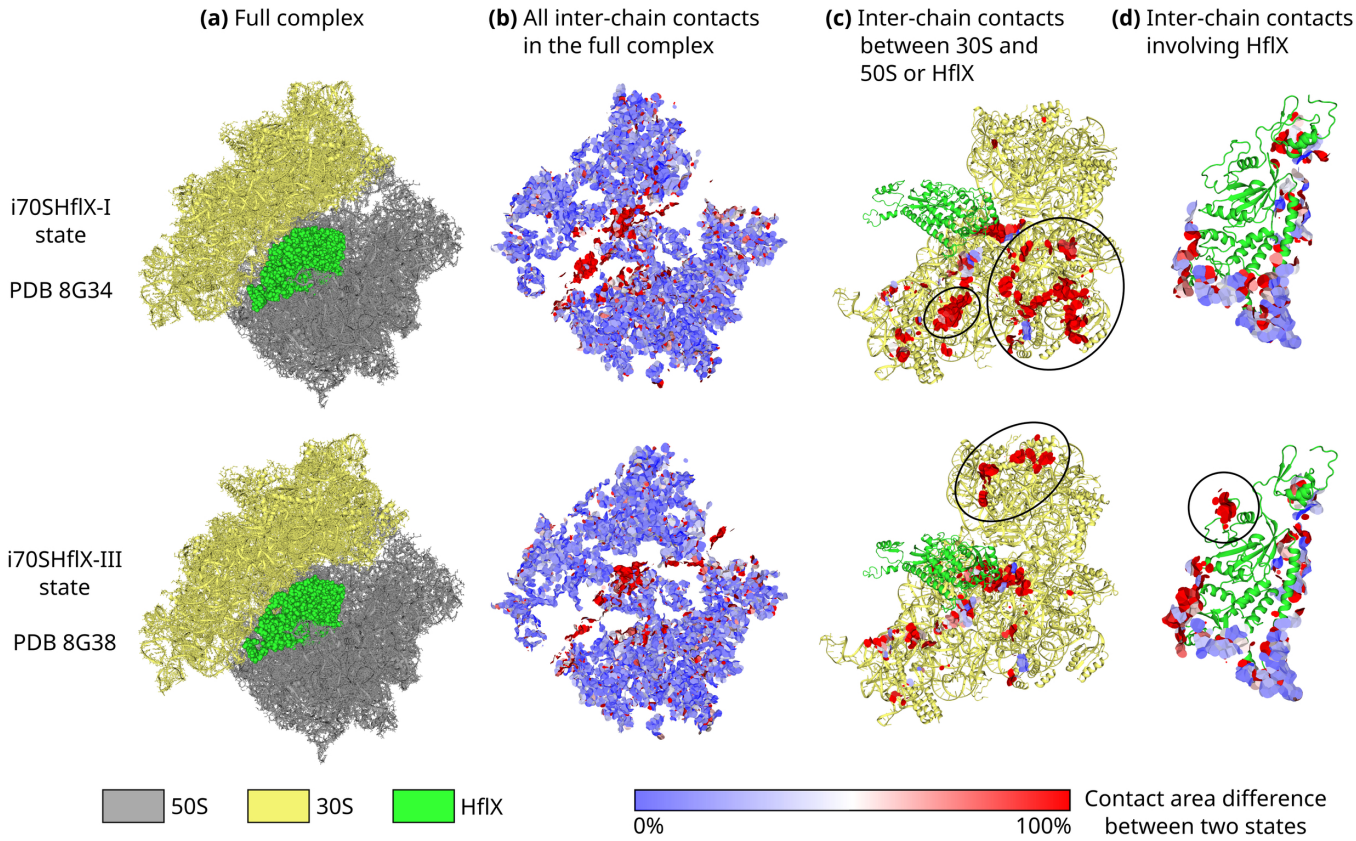

Fig. 5: Example application of Voronota-LT for studying inter-chain interface contact differences between two structural states of the 70S ribosome in complex with HflX protein that acts as a splitting factor between 50S and 30S subunits. The analysed structures were taken from the Protein Data Bank entries 8G34 and 8G38 deposited by the authors of the time-resolved cryo-EM study of HflX-mediated ribosome recycling (Bhattacharjee et al., 2023). 3D visualizations were generated using Voronota-GL.

the contracted capsid were still preserved in the extended capsid (Fig. 6).

5. Discussion and conclusions

We presented Voronota-LT, a method tailored to a very specific task — computing tessellation-based atom-atom contacts constrained inside the solvent-accessible surface of a molecular structure. Voronota-LT will not work efficiently for input sets of balls that are not similar to molecules, for example, when balls are much more densely packed or have vastly different radii. Also, unlike vanilla Voronota, Voronota-LT will work less efficiently when using a solvent probe radius much larger than 1.4 Å. However, when used for the problem that it was designed for, Voronota-LT is, to the best of our knowledge, the most efficient tool available.

Voronota-LT produces results that correlate highly with contact areas from the vanilla Voronota, but does it much faster. Moreover, because it constructs every contact directly, without needing any pre-computed global Voronoi tessellation or Delaunay triangulation, Voronota-LT can efficiently focus on only a requested subset of interactions, which is particularly useful in studies of inter-chain interfaces in biological macromolecules. Voronota-LT can construct contacts in parallel. This contact-level approach to parallelization is different from the Voronoi cell-level approach used, for example, by Voro++ (Lu et al., 2023). When focusing on inter-chain interfaces, there is no need to compute all contacts for every involved Voronoi cell. Therefore, contact-level parallelization is likely more efficient for interface studies.

Most importantly, the efficiency of Voronota-LT is coupled with the practical ability of tessellation-derived contacts to describe interactions in a comprehensive manner, fully considering the relevant structural context and providing meaningful and summarizable interaction value descriptors — contact areas. Currently there are multiple methods that use vanilla Voronota contacts, for example: CAD-score, VoroMQA, VoroIF-GNN methods used in the VoroIF-jury pipeline (Olechnovič et al., 2023); VoroCNN (Igashov et al., 2021a); S-GCN (Igashov et al., 2021b). In principle, all those methods can potentially use Voronota-LT contacts and run significantly faster. We provide Voronota-LT freely with a permissive open-source code license to enable the broad structural bioinformatics community to benefit from the tessellation-based analysis of molecular interactions without much computational resources.

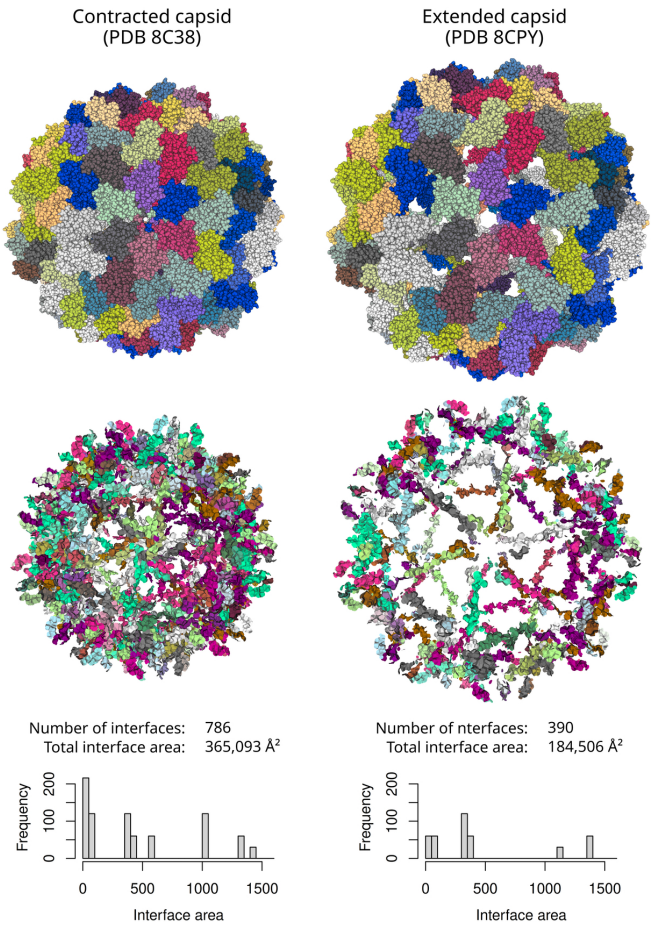

Fig. 6: Example application of Voronota-LT for studying inter-chain interfaces in contracted and expanded structures of the cowpea chlorotic mottle virus capsid. The analyzed structures were taken from the Protein Data Bank entries 8C38 and 8CPY deposited by the authors of the time-resolved cryo-EM study of viral dynamics (Harder et al., 2023). Top: all-atom visualization. Middle: visualization of inter-chain contacts constructed by Voronota-LT. Bottom: frequency of the chain-chain interface areas. 3D visualizations were generated using Voronota-GL.

Funding

Funded by the European Union

This project has received funding from the European Union under the Marie Skłodowska-Curie grant agreement No 101059190.

References

Aurenhammer, F. (1987). Power Diagrams: Properties, Algorithms and Applications. *SIAM J. Comput.*, 16(1):78–96. [doi: 10.1137/0216006].

Ben-Naim, A. (1997). Statistical potentials extracted from protein structures: Are these meaningful potentials? *The Journal of Chemical Physics*, 107(9):3698–3706. [doi: 10.1063/1.474725].

Bhattacharjee, S., Feng, X., Maji, S., Dadhwal, P., Zhang, Z., Brown, Z. P., and Frank, J. (2023). Time resolution in cryo-EM using a novel PDMS-based microfluidic chip assembly and its application to the study of HflX-mediated ribosome recycling. [doi: 10.1101/2023.01.25.525430].

Cazals, F. (2006). Revisiting the Voronoi description of protein-protein interfaces. *Protein Sci.*, 15(9):2082–2092. [doi: 10.1110/ps.062245906].

Dagum, L. and Menon, R. (1998). OpenMP: an industry standard API for shared-memory programming. *IEEE Comput. Sci. Eng.*, 5(1):46–55. [doi: 10.1109/99.660313].

Esque, J., Oguey, C., and de Brevern, A. G. (2011). Comparative analysis of threshold and tessellation methods for determining protein contacts. *J Chem Inf Model*, 51(2):493–507. [PubMed: 21226523] [doi: 10.1021/ci100195t].

Goede, A., Preissner, R., and Frömmel, C. (1997). Voronoi cell: new method for allocation of space among atoms: elimination of avoidable errors in calculation of atomic volume and density. *J. Comput. Chem.*, 18:1113–1123.

Harder, O. F., Barrass, S. V., Drabbels, M., and Lorenz, U. J. (2023). Fast viral dynamics revealed by microsecond time-resolved cryo-EM. *Nat Commun*, 14(1):5649. [PubMed: 37704664] [PubMed Central: PMC10499870] [doi: 10.1038/s41467-023-41444-x].

Igashov, I., Olechnovič, K., Kadukova, M., Venclovas, C., and Grudinin, S. (2021a). VoroCNN: Deep convolutional neural network built on 3D Voronoi tessellation of protein structures. *Bioinformatics*, page btab118. [PubMed: 33620450] [doi: 10.1093/bioinformatics/btab118].

Igashov, I., Pavlichenko, N., and Grudinin, S. (2021b). Spherical convolutions on molecular graphs for protein model quality assessment. *Mach. Learn.: Sci. Technol.*, 2(4):045005. [doi: 10.1088/2632-2153/abf856].

Imai, H., Iri, M., and Murota, K. (1985). Voronoi Diagram in the Laguerre Geometry and Its Applications. *SIAM J. Comput.*, 14(1):93–105. [doi: 10.1137/0214006].

Kim, D. S., Kim, D., and Cho, Y. (2005). Euclidean Voronoi diagrams of 3D spheres: Their construction and related problems from biochemistry. *Lect. Notes. Comput. Sc.*, 3604:255–271.

Lee, M., Sugihara, K., and Kim, D.-S. (2022). Robust Construction of Voronoi Diagrams of Spherical Balls in Three-Dimensional Space. *Comput. Aided Design*, 152:103374. [doi: 10.1016/j.cad.2022.103374].

Lensink, M. F., CAPRI-participants, and Wodak, S. J. (2023). Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment. *Proteins*, 91(12):1658–1683. [PubMed: 37905971] [doi: 10.1002/prot.26609].

Li, A. J. and Nussinov, R. (1998). A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins*, 32(1):111–127. [PubMed: 9672047].

Lu, J., Lazar, E. A., and Rycroft, C. H. (2023). An extension to Voro++ for multithreaded computation of Voronoi cells. *Computer Physics Communications*, 291:108832. [doi: 10.1016/j.cpc.2023.108832].

McConkey, B. J., Sobolev, V., and Edelman, M. (2002). Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure. *Bioinformatics*, 18(10):1365–1373. [PubMed: 12376381].

McConkey, B. J., Sobolev, V., and Edelman, M. (2003). Discrimination of native protein structures using atom-atom contact scoring. *Proc. Natl. Acad. Sci. U.S.A.*, 100(6):3215–3220. [PubMed: 12631702] [PubMed Central: PMC152272] [doi: 10.1073/pnas.0535768100].

Neira, J. L., Davis, B., Ladurner, A. G., Buckle, A. M., Gay, G. d. P., and Fersht, A. R. (1996). Towards the complete structural characterization of a protein folding pathway: the structures of the denatured, transition and native states for the association/folding of two complementary fragments of cleaved chymotrypsin inhibitor 2. Direct evidence for a nucleation-condensation mechanism. *Fold Des*, 1(3):189–208. [PubMed: 9079381] [doi: 10.1016/s1359-0278(96)00031-4].

Olechnovič, K., Valančauskas, L., Dapkūnas, J., and Venclovas, C. (2023). Prediction of protein assemblies by structure sampling followed by interface-focused scoring. *Proteins*, 91(12):1724–1733. [PubMed: 37578163] [doi: 10.1002/prot.26569].

Olechnovič, K. and Venclovas, C. (2014). Voronota: A fast and reliable tool for computing the vertices of the Voronoi diagram of atomic balls. *J. Comput. Chem.*, 35(8):672–681. [PubMed: 24523197] [doi: 10.1002/jcc.23538].

Olechnovič, K. and Venclovas, C. (2017). VoroMQA: Assessment of protein structure quality using interatomic contact areas. *Proteins*, 85(6):1131–1145. [PubMed: 28263393] [doi: 10.1002/prot.25278].

Olechnovič, K. and Venclovas, C. (2021). VoroContacts: a tool for the analysis of interatomic contacts in macromolecular structures. *Bioinformatics*, page btab448. [PubMed: 34132767] [doi: 10.1093/bioinformatics/btab448].

Olechnovič, K. and Venclovas, C. (2023). VoroIF-GNN: Voronoi tessellation-derived protein-protein interface assessment using a graph neural network. *Proteins*, 91(12):1879–1888. [PubMed: 37482904] [doi: 10.1002/prot.26554].

Poupon, A. (2004). Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. *Curr. Opin. Struct. Biol.*, 14(2):233–241. [PubMed: 15093839] [doi: 10.1016/j.sbi.2004.03.010].

Rose, Y., Duarte, J. M., Lowe, R., Segura, J., Bi, C., Bhikadiya, C., Chen, L., Rose, A. S., Bittrich, S., Burley, S. K., and Westbrook, J. D. (2021). RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive. *J Mol Biol*, 433(11):166704. [PubMed: 33186584] [PubMed Central: PMC9093041] [doi: 10.1016/j.jmb.2020.11.003].

Schweke, H., ELIXIR-participants, and Wodak, S. J. (2023). Discriminating physiological from non-physiological interfaces in structures of protein complexes: A community-wide study. *Proteomics*, 23(17):e2200323. [PubMed: 37365936] [doi: 10.1002/pmic.202200323].

Todhunter, I. (1863). *Spherical trigonometry, for the use of colleges and schools: with numerous examples*. Macmillan.

Verdonk, M. L., Ludlow, R. F., Giangreco, I., and Rathi, P. C. (2016). Protein-Ligand Informatics Force Field (PLIff): Toward a Fully Knowledge Driven "Force Field" for Biomolecular Interactions. *J. Med. Chem.*, 59(14):6891–6902. [PubMed: 27353137] [doi: 10.1021/acs.jmedchem.6b00716].

Voronoi, G. (1908). Nouvelles applications des parametres continus a la theorie des formes quadratiques. *J. Reine Angew. Math.*, 134:198–287.

wwPDB consortium (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res*, 47(D1):D520–D528. [PubMed: 30357364] [PubMed Central: PMC6324056] [doi: 10.1093/nar/gky949].