Computational characterization of biomolecular interaction specificity

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

Dr. rer. nat.

der Fakultät für

Biologie

an der

Universität Duisburg–Essen

vorgelegt von

Jean-Noël Grad

aus Saverne, Frankreich

Dezember 2021

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden in der Abteilung für Bioinformatik der Fakultät für Biologie der Universität Duisburg–Essen oder an einer anderen gleichwertigen Einrichtung durchgeführt.

1. Gutachter: Prof. Dr. Daniel Hoffmann

2. Gutachter: Prof. Dr. Thomas Schrader

3. Gutachter:

Vorsitzender des Prüfungsausschusses: Prof. Dr. Elsa Sánchez-García

Tag der mündlichen Prüfung: 3. Dezember 2021

DOI: 10.17185/duepublico/77095
URN: urn:nbn:de:hbz:465-20230329-134753-7

Contents

List of Figures

Acronyms

5-FAM	5-carboxyfluorescein
A-Raf	serine/threonine-protein kinase A-Raf
AM1	Austin Model 1
AMBER	Assisted Model Building with Energy Refinement (MD software)
APBS	Adaptive Poisson–Boltzmann Solver
ATP	adenosine triphosphate
B-Raf	serine/threonine-protein kinase B-Raf
BCC	bond charge corrections
c-Raf	RAF proto-oncogene serine/threonine-protein kinase
CAPRI	Critical Assessment of PRedicted Interactions
CHARMM	Chemistry at Harvard Macromolecular Mechanics (MD software)
COD	Crystallography Open Database
CSD	Cambridge Structural Database
CW	Cardin–Weintraub
DFT	density functional theory
DNA	deoxyribonucleic acid
DNN	distance to nearest neighbor
EC_{50}	half maximal effective concentration
EEM	electronegativity equalization method
EG	energy grid
ERK	extracellular signal-regulated kinase
ESP	electrostatic potential
FFT	fast Fourier transform
GAFF	general AMBER force field
GCP	guanidiniocarbonyl pyrrole
GLYCAM	glycan AMBER (AMBER-compatible forcefield for polysaccharides)
GPU	graphics processing unit
GROMACS	GROningen MAchine for Chemical Simulations (MD software)
HF	Hartree–Fock
LEaP	link, edit, and parm (tool to create molecular topologies)
MD	molecular dynamics
MEK	mitogen-activated protein kinase kinase
MEP	molecular electrostatic potential
MM	molecular mechanics
MOAD	Mother of All Databases
NMR	nuclear magnetic resonance
PAM	partitioning around medoids
PB	Poisson–Boltzmann
PCA	principal component analysis
PDB	Protein Data Bank

PEOE partial equalization of orbital electronegativity
PPI protein–protein interaction
PQR file format derived from the PDB file format with charges (Q) and radii (R)
QM quantum mechanics
RAF rapidly accelerated fibrosarcoma
REsP restrained electrostatic potential
RMSD root-mean-square deviation
RNA ribonucleic acid
Shh Sonic Hedgehog
SMPB size-modified Poisson–Boltzmann
TLEaP terminal LEaP (see LEaP)

Abstract

Supramolecular ligands engineered to bind biomolecules with high specificity can help elucidate biological processes by acting as biosensors, enzyme modulators or protein–protein interaction modulators. Achieving specificity requires detailed knowledge of the interaction between the ligand and target biomolecule. Obtaining this information experimentally can prove challenging, especially for flexible supramolecular ligands that do not adopt a well-defined binding pose. In these situations, computational methods can be used to model the interaction and sample the most probable binding modes.

There is a variety of techniques to model the interaction between two molecules, with a trade-off between levels of granularity and computational speed. This work is concerned with the computational evaluation of the energetic landscape of protein–polymer complexes when the electrostatic interaction is the driving force. This effort is carried out using rigid body correlation, a type of exhaustive search where the protein and ligand are treated as non-flexible entities interacting non-covalently. Suitable building blocks extracted from the polymer are used as probes to reveal areas of high affinity on the protein surface and help infer the most probable binding sites of the full-length polymer.

Inaccuracies caused by treating flexible molecules as rigid bodies are mitigated by applying the correlation on multiple conformers to generate a more realistic energetic landscape. The result of this correlation is directly comparable to crystallographic data and molecular dynamics data, and is shown to be reliable even for polymer fragments bearing a single electric charge. The method is fast and rather insensitive to the choice of charge parameters for the ligand.

Kurzfassung

Supramolekulare Liganden können so konstruiert werden, dass sie Biomoleküle mit hoher Spezifität binden und somit bei der Untersuchung biologischer Prozesse als Biosensoren, Protein–Protein-Interaktionsmodulatoren oder Enzymmodulatoren dienen. Um eine hohe Spezifität zu erreichen, benötigt man detaillierte Kenntnisse der Wechselwirkung zwischen dem Liganden und dem Zielbiomolekül. Diese Kenntnisse experimentell zu erlangen, kann sich als Herausforderung erweisen, insbesondere für flexible supramolekulare Liganden, die keine genau definierte Bindungsposition einnehmen. In diesen Fällen können numerische Methoden Abhilfe schaffen, indem man die Wechselwirkung nachbildet, und dann nach die wahrscheinlichsten Bindungsmodi sucht.

Es gibt eine Vielzahl von Techniken, um die Wechselwirkung zwischen zwei Molekülen zu modellieren, wobei ein Kompromiss zwischen Granularitätsgrad und Rechengeschwindigkeit besteht. Diese Arbeit befasst sich mit der numerischen Auswertung der Energielandschaft von Protein–Polymer-Komplexen, wenn die elektrostatische Wechselwirkung die treibende Kraft ist. Das Modell basiert auf der Verwendung der sogennanten Starrkörperkorrelation, bei der das Protein und der Ligand jeweils als starre Einheiten behandelt werden, die nicht kovalent wechselwirken. Geeignete Bausteine, die aus dem Polymer extrahiert wurden, werden als Sonden verwendet, um Bereiche mit hoher Affinität auf der Proteinoberfläche zu entdecken und die wahrscheinlichsten Bindungsstellen des Polymers voller Länge abzuleiten.

Ungenauigkeiten, die durch die Behandlung flexibler Moleküle als starre Körper verursacht werden, werden durch Anwenden der Korrelation auf mehrere Konformere verringert, um eine realistischere Energielandschaft zu erzeugen. Das Ergebnis dieser Korrelation ist direkt mit kristallographischen Daten und molekulardynamischen Daten vergleichbar und hat sich auch für Polymerfragmente mit einer einzigen elektrischen Ladung als zuverlässig erwiesen. Das Verfahren ist schnell und ziemlich unempfindlich gegenüber der Wahl der Ladungsparameter für den Ligand.

Acknowledgments

I would like to thank my supervisor Daniel Hoffmann for his continuous support during my PhD. You offered critical input on my computational methods and statistical analysis tools, helped me overcome obstacles in my research projects, and created an extremely motivating work environment for me and my colleagues. I really enjoyed our long discussions on the scientific literature of molecular dynamics methods and protein–ligand correlation algorithms. I am extremely grateful to my co-supervisor Carsten Schmuck for his input as an experimentalist and his work at the CRC 1093. My thanks also go to Thomas Schrader, Elsa Sánchez-García, Shirley Knauer and Andrea Vortkamp for fruitful discussions about their research projects.

I would like to thank my colleagues from the Bioinformatics and Computational Biophysics group and from the CRC 1093, in particular: Ludwig Ohl for our rich conversations about physics and chemistry topics, Rocio Rebollido-Rios for your help with molecular dynamics and quantum mechanical calculations, Martin Ehlers and David Bier for our exchanges on molecular modeling and crystallography, Sumit Mittal and Alba Gigante for your input on my modeling work with 14-3-3, and Christoph Wilms for kindly sharing your Python expertise. Special thanks to Anja Lange for your reactivity as a sysadmin and for maintaining a complex IT infrastructure, Claudia Wilmes for your help with administrative tasks, and Lydia Didt-Koziel for organizing the CRC 1093 conferences and symposia.

My gratitude also goes to Mike Nemec, Bettina Budeus, Daniel Habermann, Simo Kitanovski, Niklas Tötsch, Irina Curuira, Farnoush Farahpour, Mohammadkarim Saeedghalati and Marius Welzel for the scientific exchanges and the friendly atmosphere at the office, and to all the CRC 1093 colleagues for the amazing collaborative work.

Finally, I would like to thank my colleagues at the Institute for Computational Physics of Stuttgart, especially Kai Szuttor for kindly accepting to proofread my thesis, as well as my family for their support during my time as a PhD student.

Chapter 1
Modeling molecular recognition

Elucidation of the physico-chemical mechanism for molecular recognition plays a critical role in rational ligand design, where knowledge of the binding mode gained from a combination of experimental data and computational methods is used to develop ligands with enhanced affinity and specificity for a target (Tollenaere 1996). Although more than a century old, rational ligand design is a constantly evolving field of research, notably in supramolecular chemistry where techniques originally designed for small and relatively rigid molecules aren't well-suited to the investigation of long and flexible polymers (Schneider 2009).

The "lock-and-key" principle was the first attempt at modeling molecular recognition, relying on geometric complementarity to explain enzyme–substrate specificity (Fischer 1894). The purely rigid and sterical nature of this model would later evolve to capture protein flexibility and long-range interactions (Schneider 2015; Csermely et al. 2010). Two well-known examples are the "induced fit" paradigm (Koshland 1958) and the "conformational selection" paradigm (Csermely et al. 2010), created to account for enzyme plasticity at the catalytic site. Although the difference between these two paradigms is still a matter of debate to this day (Csermely et al. 2010; Gianni et al. 2014; Chakraborty and Di Cera 2017; Di Cera 2020), other developments reduced the relevance of the lock-and-key paradigm, such as conformational proofreading (Cramer and Freist 1987), allosteric effects (Lavigne and Anslyn 2001; Schneider 2015) and hydrophobic effects (Schneider 2015), to name a few examples. In spite of its simplicity, the lock-and-key metaphor can still be used in biology as a form of abstraction (Cramer 2007), for example when describing computational chemistry algorithms (Morrison et al. 2006).

Rational ligand design often relies on protein crystallography, an experimental technique in structural biology that generates macromolecular models of protein–ligand complexes with near-atomic precision (Verlinde and Hol 1994; Anderson 2003). Since this technique can only resolve static protein–ligand systems with sufficient accuracy to produce interpretable data, it is of limited use for flexible protein–polymer systems, although polymer fragments can sometimes be co-crystallized with their target protein to reveal high-affinity binding sites. The structural information gained via this approach offers an incomplete picture of the binding mode, as polymers typically interact with their target over a large surface area. Low-affinity binding sites are also difficult to detect with this technique, even though they contribute to the total binding affinity of the protein–polymer complex in a non-linear manner, which is known as the "cooperative effect" (Di Cera 1998; Biedermann and Schneider 2016).

Computer-aided ligand design provides an alternative in the form of a molecular mechanics (MM) model of the interaction. In general terms, modeling a protein–polymer system requires the generation of a topology for the protein and polymer, and the formulation of an energy function—the Hamiltonian—that depends on geometrical parameters. The evaluation of the Hamiltonian for a collection of geometrical parameters yields an energetic landscape, which can be used to test hypotheses on the protein–polymer system. For example, calculating the derivative of this energy function gives the direction of the nearest local minimum or saddle point, from which the ligand position and geometry can be iteratively optimized until a binding site is found. This strategy is usually known as an optimized search, and can be carried out with deterministic and stochastic algorithms.

Optimized search is the method of choice for protein–ligand systems with a single binding site, however in the context of protein–polymer interfaces, multiple binding sites contribute to the stability of the complex. For these systems it is desirable to sample the energetic landscape uniformly using exhaustive search algorithms, such as correlation-based algorithms. This approach is typically slower and more memory-intensive than an optimized search, and relies on a coarse description of the interaction and molecular geometry, but detects all local minima in that model within a specified tolerance. Under favorable conditions, the results from such a calculation can be used as input parameters of an optimized search to generate a higher-resolution model of the interaction.

1.1 Rigid body correlation: An overview

1.1.1 Early work on correlation

Greer and Bush formulated a framework for the encoding of physico-chemical properties into grid-based representations of biomolecules and the evaluation of correlation functions using these representations to detect complementarity sites (Greer and Bush 1978). Multiple algorithms were developed in the following years to exploit this framework but were limited to either two-dimensional exhaustive searches via the cross-correlation of protein and ligand slices (Zielenkiewicz and Rabczenko 1984; Zielenkiewicz and Rabczenko 1985), or to a heuristic three-dimensional search (Kuntz et al. 1982). GRID was one of the first methods to propose an exhaustive three-dimensional sampling of the volume around a protein, using a spherical probe to represent a chemical group made of one to four atoms (Goodford 1985). This simplification ignored any information about the ligand structure and orientation, but the peaks in the resulting grid could be used to manually place a ligand in the putative binding site. The energetic model took Coulomb and van der Waals interactions into account.

AUTODOCK 1 (Goodsell and Olson 1990) used the GRID methodology to pre-compute energetic grids for all mono-atomic probes extracted from a ligand and store the grids in memory as lookup tables. These grids were computed in a small region of the putative binding site to reduce their memory footprint. Evaluating the energy of multiple ligand conformations was achieved via an inexpensive linear interpolation of the grid points around the position of each atom in the ligand. This strategy enabled the optimization of a ligand conformation by simulated annealing (Kirkpatrick et al. 1983; Černý 1985), a type of heuristic search.

SOFTDOCK introduced polyatomic probes and relied on the exhaustive six-dimensional sampling of the volume around a protein, i.e. sampling of the three translational and three rotational degrees of freedom of the probe (Jiang and Kim 1991). The energetic model was simplified to consider a binary response—favorable/unfavorable—between three interactions: electrostatics, hydrogen bonding and hydrophobicity. The energy grid was calculated as the sum of favorable interactions on the contact surface between the protein and ligand. In an effort to make the calculation computationally tractable, it was carried out in two steps: an exhaustive six-dimensional sampling using a coarse grid resolution and a small number of rotations to detect putative binding sites quickly, followed by an exhaustive six-dimensional refinement of each candidate binding site.

1.1.2 The fast Fourier transform breakthrough

So far, software relying on exhaustive three-dimensional (only translations) or six-dimensional (translations and rotations) correlation algorithms were limited by the computational power and memory of the hardware. However, the correlation of three-dimensional matrices in real space is equivalent to a matrix multiplication in the frequency domain according to the cross-correlation theorem (Stockham 1966), hence any three-dimensional translation correlation can be evaluated using the fast Fourier transform (FFT) algorithm (Cooley and Tukey 1965) to reduce the computational complexity from $\mathcal{O}(N^6)$ to $\mathcal{O}(N^3 \log(N^3))$. This strategy made calculations involving larger probes computationally amenable, such as small proteins (Katchalski-Katzir et al. 1992; Gabb et al. 1997). Similarly, energetic models could afford to be more elaborate, and therefore more realistic, as long as interactions could be modeled in a linear form.

MolFit (Katchalski-Katzir et al. 1992) was the first software to take advantage of the FFT to speed up the exhaustive translational correlation of two proteins with atomistic resolution. FFT acceleration quickly became the norm for rigid body correlation, even for low-resolution correlation (Vakser and Aflalo 1994; Vakser 1995; Vakser and Nikiforovich 1995), which lead to the GRAMM (Vakser 1997) and GRAMM-X (Tovchigrechko and Vakser 2005; Tovchigrechko and Vakser 2006) software. A two-stage approach was developed where a rigid body exhaustive search produces putative binding sites on the target protein and putative orientations of the partner protein at these sites, followed by a flexible body refinement of the most promising candidates, creating a link between the fields of shape correlation and molecular docking. This approach was implemented by Ckordo (Meyer et al. 1996; Heuser et al. 2005), FTDock (Gabb et al. 1997), 3D-Dock (Smith and Sternberg 2003) and RDOCK (Li et al. 2003). More in-depth details can be found in contemporaneous reviews (Halperin et al. 2002; Eisenstein and Katchalski-Katzir 2004; Vajda and Camacho 2004; Moreira et al. 2010), as molecular docking is outside the scope of this work.

Further research in FFT algorithms lead to the spherical polar Fourier transform (Ritchie and Kemp 2000; Kovacs and Wriggers 2002; Kovacs et al. 2003) to accelerate the exhaustive six-dimensional correlation in Hex (Ritchie and Kemp 2000; Ritchie et al. 2008; Ritchie and Grudinin 2016) and FRODOCK (Garzon et al. 2009). A GPU-accelerated FFT algorithm is used in Hex (Ritchie and Venkatraman 2010) and PIPER (Sukhwani and Herbordt 2010).

1.1.3 Diversification of the energetic model

While the original implementation of MolFit was based on geometric matching, the energetic model later included the hydrophobic effect (Vakser and Aflalo 1994) and the electrostatic interaction using the linearized Poisson–Boltzmann (PB) equation (Heifetz et al. 2002). An early version of GRAMM (Vakser 1995) adapted the geometric matching strategy for large and partially disordered protein structures with low atomic resolution, later evolving into GRAMM-X (Tovchigrechko and Vakser 2005) for high-resolution protein structures, using a modified Lennard-Jones potential for improved accuracy. FTDock (Gabb et al. 1997) added a simple Coulomb formula with a variable dielectric to treat the electrostatic interaction, while DOT (Mandell et al. 2001) and CS-Map (Dennis et al. 2002; Kortvelyesi et al. 2003) used the linearized PB equation. FTDock later evolved into ZDock (Chen and Weng 2002) through the addition of a desolvation term in the energetic model; the shape complementarity term in the energetic model was ultimately replaced with pairwise potentials (Mintseris et al. 2007). PIPER (Kozakov et al. 2006) also uses a combination of shape complementarity, electrostatics and pairwise potentials, while Ckordo (Meyer et al. 1996) uses hydrogen bonding and shape complementarity.

The pair potentials method expresses inter-atomic potentials in the form $\sum_{i,j\neq i} a_i a_j / r_{i,j}^n$ with $a_i a_j$ the interaction prefactors for atoms i and j and $r_{i,j}^n$ the separating distance raised to the power n (Harrison et al. 1994). These convolutions can be evaluated at a low computational cost using FFT, although some potentials are poor candidates. In particular, the electrostatic potential in a non-uniform dielectric cannot be rewritten as a pair potential. In that case, the electrostatic potential is either approximated with a uniform dielectric, or calculated from the PB equation in a numerical solver.

The statistical pair potentials method is an alternative strategy that relies on datasets of crystallographic coordinates of protein–protein complexes to estimate potentials in the form $\sum_{i,j\neq i} e_i e_j$ within a distance threshold (Miyazawa and Jernigan 1985). The potential mean force between atom types can be evaluated from the frequency of pairwise contacts in the dataset using the Boltzmann distribution (Zhang et al. 1997; Moont et al. 1999), as implemented in FTDock (Moont et al. 1999), ZDock (Chen and Weng 2002; Mintseris et al. 2007) and CS-Map (Dennis et al. 2002; Kortvelyesi et al. 2003). The pairwise potentials can be simplified further by reducing the set of interactions e_i to a small ensemble of linear combinations, as implemented in PIPER (Kozakov et al. 2006) and ZDock (Mintseris et al. 2007).

1.1.4 The case of small ligand correlation

The development of the first FFT-accelerated correlation algorithm enabled protein–protein docking in a matter of hours (Katchalski-Katzir et al. 1992) and profoundly changed the range of applications of exhaustive scanning, making it a viable alternative to established stochastic search algorithms. By 2010, FFT-based protein–protein docking algorithms became as popular as stochastic algorithms (Moreira et al. 2010) in the CAPRI challenge (Janin et al. 2003).

This success overshadowed the original protein–ligand correlation methods in GRID and SOFTDOCK. Research on protein–ligand scanning was neglected until the release of CS-Map (Dennis et al. 2002) for protein–solvent mapping. The project eventually evolved into FTMap

(Brenke et al. 2009), a tool to infer the binding site of large inhibitors from the mapping of small solvent molecules. The tool was later extended to support the mapping of custom drug-like compounds (Ngan et al. 2012). This was followed by the release of an early version of Epitopsy tailored for protein–ligand scanning (Wilms 2013) based on the original FTDock algorithm (Gabb et al. 1997), but using the non-linear PB equation for a more accurate description of charge–charge interactions.

The fragment-based scanning methodology originally explored in GRID (Goodford 1985), SOFTDOCK (Jiang and Kim 1991) and FTMap (Brenke et al. 2009) eventually found applications in protein–polymer interaction modeling. Using chemically relevant fragments of a polymer, for example its constituent monomers, one could "map" hot-spots on the protein surface with a protein–ligand correlation tool and use that information to formulate an hypothesis on the putative protein–polymer binding mechanism. This strategy was originally investigated with sulfated polysaccharide fragments as the molecular probe by Epitopsy (Wilms 2013) and PIPER (Mottarella et al. 2014). The present work builds upon and extends Epitopsy for protein–ligand correlation and validates its application in protein–polymer research, using heparin (Grad et al. 2018) and a semi-synthetic polymer (Gigante et al. 2019) as highlights.

1.1.5 Epitopsy: A general purpose protein–ligand correlation tool

1.1.5.1 Hamiltonian

Soon after the release of the first version of Molfit (Katchalski-Katzir et al. 1992), a derivation of the Helmholtz free energy of binding in the grand canonical ensemble was formulated in the pair potentials method (Harrison et al. 1994), using an approximation of the electrostatic potential that ignores the change of dielectric permittivity at the protein–solvent interface. The authors suggested using an electrostatics solver to calculate the electrostatic potential of the protein on a grid with the same dimensions as the shape grid. This idea was implemented simultaneously by MolFit (Heifetz et al. 2002), DOT (Mandell et al. 2001) and CS-Map (Dennis et al. 2002) using the linearized PB equation.

A derivation of the Gibbs free energy of binding was formulated in prior work on Epitopsy (Wilms 2013), with emphasis on the PB equation. This formulation is reproduced and extended in Appendix A on page 86, using a different notation and providing a direct link between discretized quantities and continuum quantities. The discretization and correlation algorithm are detailed in greater depth in Work 1 (Grad et al. 2018, equations 1 to 13).

1.1.5.2 Software implementation

Epitopsy is an open-source Python (Oliphant 2007) package licensed under the GNU LGPL v3 (Free Software Foundation 2007). Calculations are carried out with the following scientific computing libraries: BioPython for operations on molecular structures (Cock et al. 2009), NumPy for operations on arrays (Oliphant 2006; Millman and Aivazis 2011), ANFFT for FFT (Collette 2010) and *scikit-learn* for machine learning (Pedregosa et al. 2011). Performance-critical code is written in Cython (Behnel et al. 2011).

1.1.5.3 Input files

Epitopsy requires two types of inputs: molecular topologies and electrostatic potential grids. A molecular topology consists of an all-atom three-dimensional molecular structure with atomic radii and partial charges information. These parameters are typically obtained from MM force fields using the tool PDB2PQR (Dolinsky et al. 2004; Dolinsky et al. 2007). The protein and ligand structures need to be provided in the PQR file format, which is a derivative of the PDB file format with extra fields to store the partial charges and atomic radii.

The non-linear PB equation is solved by APBS (Baker et al. 2001; Jurrus et al. 2018) using input files generated by PDB2PQR. The electrostatic potential grids of biomolecules and the corresponding boundary grids must be provided to Epitopsy in the OpenDX file format, which is the default output of the APBS software.

1.1.5.4 Obtaining suitable partial charges

Ab initio methods

Quantum chemistry packages can derive partial charges from the electrostatic potential of small molecules relatively inexpensively. Partial charges can also be obtained from MM force fields derived from *ab initio* calculations, such as the AMBER force field (Cieplak et al. 1995).

Semi-empirical methods

The Austin Model 1 (AM1) (Dewar et al. 1985) is a fast but inaccurate semi-empirical method for deriving partial charges (Jakalian et al. 2002). The AM1-BCC method uses the AM1 method to derive approximate charges and refines them via bond charge corrections (BCC) to reproduce HF/6-31G*//HF/6-31G* charges[1] with good accuracy (Jakalian et al. 2000; Jakalian et al. 2002). This approach is suitable when working with large libraries of drug-like compounds, for which a systematic HF/6-31G*//HF/6-31G* charge derivation would be too expensive. This approach is used in the GAFF force field (Wang et al. 2004; Jämbeck and Lyubartsev 2014).

Empirical methods

The partial equalization of orbital electronegativity (PEOE) method (Gasteiger and Marsili 1978; Gasteiger and Marsili 1980) assigns charges based on electronegativities and bond information. The calculation is fast at the expense of accuracy (Bikadi and Hazai 2009; Santos-Martins et al. 2014), sometimes requiring manual corrections for protonable groups (Huey et al. 2007; Morris et al. 2009). It is the default method in Open Babel (O'Boyle et al. 2011). The empirical electronegativity equalization method (EEM) (Jirouškovà et al. 2009) provides a more accurate alternative to the PEOE method. Originally derived from a semi-empirical method (Mortier et al. 1985; Mortier et al. 1986), the empirical version is parameterized using datasets of ligands for which *ab initio* partial charges have been calculated. Parameter sets are available for non-standard ligands (Geidl et al. 2015; Raček et al. 2016).

[1] *Ab initio* calculations at two levels of theory are written {single-point energy//geometry}. HF stands for the Hartree–Fock method (Hartree and Hartree 1935). The Pople basis set 6-31G* describes the number of Gaussian functions used per electron in the system (Ditchfield et al. 1971).

1.1.5.5 Rotation function

Rigid body correlation in six-dimensional space requires uniform sampling methods for both translations and rotations. While partitioning translations uniformly in $\mathbb{R}^3$ is straightforward, generating an isotropic distribution of rotations of arbitrary size in $\mathbb{R}^3$ is a long-standing problem in mathematical modeling. It is especially crucial in rigid body correlation, as any bias would lead to the generation of preferential orientations, making the evaluation of shape and electrostatic complementarity sensitive to the initial orientation of the input ligand structure.

Trivial solutions to this problem can be derived from molecular symmetry: the non-reflective symmetry operations of the T_d, O_h and I_h groups generate uniform and orientation-preserving rotation sets of size 12, 24, and 60 respectively; reflections and inversions are excluded as they do not preserve chirality. To generate pseudo-uniform sets of angular states of arbitrary size, the axis-angle method (Miles 1965) can be used, which reduces the problem to finding a pseudo-uniform distribution of points on the surface of a sphere around which the probe can spin. Fibonacci generative spirals (Hannay and Nye 2004; Swinbank and James Purser 2006; González 2010) were used throughout this work to sample the rotation axis, while the spin was sampled from a uniform distribution. A summary of existing sampling methods is available in Appendix B on page 88.

1.2 Molecular mechanics

Three-dimensional models of proteins and ligands suitable for rigid body correlation and molecular dynamics can be readily obtained from structural biology databases. Familiarity with the experimental techniques can help identify sources of bias in the structural data, which need to be corrected during the pre-processing stage.

Generating a topology suitable for molecular dynamics (MD) simulations is straightforward for proteins, which have default sets of parameters in most MM force fields. Non-natural polymers however need to be parameterized.

The following sections give an overview of experimental sources of protein and ligand models. Since this work only relied on protein structures sourced from X-ray diffraction data, other experimental techniques are not emphasized for the sake of brevity. Tools used for topology generation and force field parameterization are also presented.

1.2.1 Modeling proteins

1.2.1.1 Obtaining a model from structural biology databases

Protein structures can be determined experimentally using a variety of methods. Examples of well-established techniques are X-ray crystallography (Parker 2003), nuclear magnetic resonance (NMR) spectroscopy (Wüthrich 1995), neutron crystallography (O'Dell et al. 2016) and electron diffraction (Nannenga 2020). These experimental structures are typically deposited on the Protein Data Bank (PDB) (Bernstein et al. 1977; Berman et al. 2000).

1.2.1.2 Structure preparation

Background on protein crystallography

Crystallography provides a mean of determining the three-dimensional structure of a protein based on the X-ray diffraction pattern of a single crystal of that protein (Parker 2003). A single crystal is obtained by gradually concentrating an already saturated solution of the protein until nucleation occurs (Russo Krauss et al. 2013; McPherson 2017). Various organic molecules (Sauter et al. 1999) and inorganic salts (Hegde et al. 2017) are added to the solution as adjutants to trigger nucleation and promote crystal growth (Russo Krauss et al. 2013). The protein crystal is then flash-cooled with liquid nitrogen and illuminated by an X-ray source to produce a diffraction pattern. The intensity of the diffraction peaks gives information on the amplitude of the waves that form the pattern, but no phase information can be recovered. This issue is called the "phase problem" (Taylor 2010).

To construct the electron density of the unit cell, the phases of the waves are obtained from another source, usually from already solved structures with similar amino acid sequence and three-dimensional structure, or from the pooling of multiple X-ray diffraction patterns from the same protein co-crystallized with different heavy metals (Taylor 2010). The electron density map is obtained by a Fourier transform of the amplitudes and phases of the scattered waves. An initial molecular model of the protein is fitted against the electron density and the discrepancies are minimized through several rounds of refinement until a satisfactory model is obtained (Smyth and Martin 2000).

Structure quality and interpretability

Crystal structures suffer from different sources of bias inherent to the techniques used in their obtention (Borek et al. 2003; Wlodawer et al. 2008). For example, crystals are diffracted at cryogenic temperatures to reduce protein mobility, with the potential to immobilize the protein conformation into an energy minimum not observed at room temperature (Fraser et al. 2011). In addition, low occupancy side-chain rotamers and alternative inhibitor binding poses might not be clearly visible (Lang et al. 2014).

Several features of the solid state are physiologically irrelevant and can potentially lead to misinterpretation. In particular, a crystal lattice is stabilized by non-covalent interactions between the protein in the unit cell and its periodic images. Identifying the physiologically relevant protein quaternary structure can sometimes prove challenging (Elez et al. 2020). X-ray structures also feature non-physiological ligands obtained from the crystallization medium which can locally perturb the protein structure (Handing et al. 2018; Ledvina et al. 1996): pH buffers (sulfate, citrate), anti-freeze agents (polyethylene glycol), co-solvents (dioxane, dimethyl sulfoxide), inorganic salts and metallic ions, to cite a few examples.

The protein may have also been subjected to mutagenesis to facilitate crystallization. These engineered proteins often deviate from their wild type variants by point mutations in solvent-exposed residues to help solubility (Derewenda 2010) or reduce entropy (Cooper et al. 2007; Derewenda 2010). Sometimes multiple residues are deleted in flexible regions to locally reduce disorder, e.g. termini trimming and loop contraction (Derewenda 2010).

Pre-processing
Experimentally-determined protein structures cannot be used directly as input to bioinformatics tools. Measurement uncertainties and perturbations must be addressed manually, namely:

- missing residues or side-chains in disordered regions, such as flexible loops
- missing hydrogen atoms; requires careful examination in the case of histidines, which have three protonation states: $N\delta$, $N\varepsilon$, or both $N\delta$ and $N\varepsilon$
- artificial insertions/deletions/mutations designed to reduce protein flexibility, facilitate purification (polyhistidine-tag) or promote crystallization
- post-translational modifications (phosphorylation, glycosylation) or lack thereof
- alternative side chain conformations
- extra ligands, salts and co-solvents
- free or capped *N*- and *C*-termini.

These details must be carefully reviewed based on the quality of the experimental data (Kleywegt 2000) and known post-translational modifications of the protein. Most operations involving the restoration of missing atoms or residues can be easily achieved with homology modeling software. Multiple validation tools are available to assess the quality of the molecular model (Laskowski 2003).

1.2.1.3 Topology preparation

For rigid body correlation, this step consists in adding van der Waals radii and atomic partial charges to the PDB file to generate a PQR file. Proteins consisting of standard amino acids and crystallographic water molecules can be directly processed by PDB2PQR (Dolinsky et al. 2004; Dolinsky et al. 2007), which uses either the CHARMM (MacKerell et al. 1998) or the AMBER (Wang et al. 2000) force field. For proteins with metallic centers, the metallic ions need to be manually added back into the PQR file, with van der Waals radii obtained from the literature and partial charges corresponding to their oxidation state.

The procedure is more complex for chemically functionalized proteins. The protein can be treated by the AMBER force field, while the functionalizing group is treated with a specialized force field, such as GLYCAM for saccharide conjugates (Woods et al. 1995) or AMBER-DYES (Graen et al. 2014) for fluorescent dyes. Other types of derivation require the use of ad hoc force field modification files from the literature.

The AmberTools suite of programs provides the LEaP tool to generate topologies from the terminal (TLEaP). The protein structure can be loaded in LEaP and then conjugated or loaded as an already-conjugated structure. The resulting MM topology can be parsed to extract the protein van der Waals radii and partial charges, and used for MD simulations.

1.2.1.4 Conformational sampling

Conformational sampling can be carried out by MD simulations if a suitable MM topology was generated at the previous step. The resulting trajectories are then clustered to extract probability-weighted conformer ensembles. Alternatively, a collection of crystal structures of the protein can be used.

1.2.2 Modeling ligands

1.2.2.1 Obtaining a model

Crystal structure of the solid phase
The CSD (Allen 2002) and COD (Gražulis et al. 2012) provide X-ray structures of ligands in the solid phase. These structures are useful to get an initial geometry for the ligand, however they do not necessarily reflect the ligand conformation(s) in solution.

Crystal structure of the bound state
The PDB (Bernstein et al. 1977; Berman et al. 2000) provides a comprehensive dataset of protein–ligand complexes. The PDBbind (Wang et al. 2005) and MOAD (Hu et al. 2005) are manually curated databases of high-resolution complexes with binding affinity data. The bound-state conformation of a ligand can be extracted from these structures.

Molecular mechanics model with experimental geometric parameters
Many natural products are readily available in parameterized form. For all-atom models, natural amino acids and nucleic acids are available in AMBER, lipids and cholesterol in LIPID (Skjevik et al. 2012; Dickson et al. 2014), saccharides and glycosaminoglycans in GLYCAM (Woods et al. 1995; Kirschner et al. 2008; Singh et al. 2016). Scaffolding of the full molecule can be achieved in AmberTools or online for GLYCAM and LIPID. The only necessary pieces of information are the experimentally-determined dihedral angles between the sub-units of the polymer. For example, dihedral angles of the glycosidic bonds in polysaccharides can be obtained from X-ray fiber diffraction.

***Ab initio* calculation**
Quantum chemistry packages can optimize a ligand geometry and generate partial charges. Despite the computational burden associated with these calculations, a reasonable trade-off between computational time and accuracy can be achieved using a vacuum optimization with Hartree–Fock (HF) or density functional theory (DFT). The HF/6-31G* level of theory for example was used to parameterize early AMBER force fields (Cornell et al. 1995).

1.2.2.2 Topology preparation

For peptides, the PDB2PQR tool will generate adequate partial charges and atomic radii. For ligands obtained from specialized force fields, the charges and radii can be extracted from the MM topology. For drug-like compounds, a literature search might reveal readily-available parameterized files. For other, non-standard ligands, charges need to be derived with the methods previously described in section 1.1.5.4 “Obtaining suitable partial charges”, while radii can usually be obtained by analogy with similar molecules with known radii.

If MD simulations are required for these non-standard ligands, they will have to be fully parameterized. For small molecules, the GAFF (Wang et al. 2004) tool provides a framework to derive bonded and non-bonded parameters by analogy with AMBER parameters. Alternatively, the PRODRG (van Aalten et al. 1996; Schüttelkopf and van Aalten 2004) tool with charge correction (Medeiros et al. 2013) can generate AMBER and GROMACS topologies. For larger

molecules, a fragment-based approach might be required. This approach consists in splitting the molecule into physiologically relevant building blocks, typically at amide and ester bonds, and deriving suitable force field parameters for each block using *ab initio* methods. The protocol is described in more length in section 1.2.3.2 “Force field parametrization”.

1.2.2.3 Conformational sampling

For small ligands, GPU-accelerated MD simulations are generally fast enough to produce microsecond trajectories in a matter of days. The trajectory can then be clustered to extract representative conformers. For polymer fragments, fiber diffraction and NMR experiments can provide experimental ranges for dihedrals connecting the polymer building blocks.

1.2.3 Molecular Dynamics

1.2.3.1 Defining the Hamiltonian

The role of a MM model is to approximate bonded and non-bonded interactions occurring at the molecular level. Non-bonded interactions are quite diverse: hydrogen bonds, salt bridges, ion–dipole and dipole–dipole forces, forces involving induced dipoles, and other van der Waals forces. In chiral organic molecules, these non-bonded interactions tend to exhibit some degree of directionality, including the charge–charge interaction, as observed for salt-bridges in solid (Barlow and Thornton 1983; Kumar and Nussinov 2002; Donald et al. 2011) and solvated (Huerta-Viga et al. 2015) systems.

The AMBER Hamiltonian is given by equation 1.1. The force is obtained by deriving the Hamiltonian with respect to the distance r. The equation can be divided into the following terms (Cornell et al. 1995; Salomon-Ferrer et al. 2013; Betz and Walker 2015):

- bonds lengths: inter-nuclear separations ℓ_i are constrained near their equilibrium value $\ell_{i,0}$ with restoring forces k_i^b using a harmonic oscillator, which for small deviations from $\ell_{i,0}$ is a reasonable approximation of the Morse potential
- bond angles: angles θ_i are constrained near their equilibrium value $\theta_{i,0}$ with restoring forces k_i^a using a harmonic oscillator
- dihedral angles: angles ϕ_i are constrained to their local minima $\phi_{i,n,0}$ with energy barriers $V_{i,n}$ and periodicity n using Fourier series
- non-bonded interactions: short-range interactions are modeled using a 6-12 Lennard-Jones potential while electrostatic forces are modeled using a Coulomb potential; for pairs of atoms separated by 3 bonds or less this calculation is not performed since their contribution to the Hamiltonian is already accounted for by the bonded terms.

$$
\begin{aligned}
V(r^N) = &\sum_{i}^{\text{bonds}} k_i^b(\ell_i - \ell_{i,0})^2 + \sum_{i}^{\text{angles}} k_i^a(\theta_i - \theta_{i,0})^2 + \sum_{i}^{\text{torsions}} \sum_{n} \frac{1}{2} V_{i,n}[1 + \cos(n\phi_i - \phi_{i,n,0})] \\
&+ \sum_{i<j} \left\{ \left[\frac{\sigma_{ij}^A}{r_{ij}^{12}} - \frac{\sigma_{ij}^B}{r_{ij}^6} \right] + \frac{q_i q_j}{\varepsilon r_{ij}} \right\}
\end{aligned}
\tag{1.1}
$$

1.2.3.2 Force field parameterization

The AMBER99SB force field provides parameters for amino acids, nucleic acids, water and ions. Non-natural molecules can be represented with specialized force fields, such as GAFF for drug-like molecules (Wang et al. 2004). Non-natural residues connected to amino acids via covalent bonds need to be parameterized by following the procedure of the original AMBER force field (Cornell et al. 1995; Cieplak et al. 1995).

If the set of atom types for natural residues in AMBER99SB is transferable to the non-natural residues to be parameterized, the Lennard-Jones parameters can be obtained by analogy (Weiner et al. 1984). The Coulomb interaction however requires the derivation of partial charges from quantum mechanics (QM) calculations (more details in Appendix C on page 93). Using a point charge model significantly simplifies the treatment of electrostatic interactions between particles, however this approximation is inadequate for highly polar molecules such as water. This issue is remediated using one of the numerous modified water models: the SPC/E model applies a polarization correction (Berendsen et al. 1987), the TIP4P model uses one extra partial charge in between the two hydrogens (Jorgensen et al. 1983), and TIP5P uses two extra partial charges for the electron pairs on the oxygen atom (Mahoney and Jorgensen 2000).

Bonded parameters are obtained through the minimization of the error between QM and MM energy for an ensemble of molecular conformations, for example via the program *paramfit* (Betz and Walker 2015). Dihedral bonds are fitted last, as their role is to encode short-range intramolecular Lennard-Jones and Coulomb interactions, as well as interactions not explicitly accounted for in the Hamiltonian, such as aromaticity. There are two types of dihedrals: proper and improper. Proper dihedrals are assigned to 4 atoms connected linearly (A–B–C–D) and represent the angle between the planes formed by the triangles ABC and BCD. They are used to model the energetic profile of the rotation barrier around the bond B–C. The Fourier sum usually contains 1 to 4 terms with varying strength, frequency and phase, depending on the complexity of the energetic landscape. Improper dihedrals are assigned to 4 atoms connected in a planar star-shaped configuration (A–B(=C)–D). They are used to model the energy penalty associated with the deformation of the plane. This configuration is used for delocalized π systems, such as aromatic cycles, amides and carbonyl groups. The Fourier sum usually consists of a single term with a high energy barrier.

Chapter 2
Validation of Epitopsy against MD simulations

Most animals with bilateral symmetry express Hedgehog morphogens (Ingham et al. 2011), which are key components of the Hedgehog signaling pathway (Robbins et al. 2012). Disruption of the pathway can lead to abnormal limb development in mouse embryos (Riddle et al. 1993; Pathi et al. 2001; Li et al. 2006) and basal cell carcinomas in humans (Daya-Grosjean and Couvé-Privat 2005). Interactions between Sonic Hedgehog (Shh) and glycosaminoglycans are an active field of study (Zhang et al. 2007; Manikowski et al. 2019), and are being investigated as potential targets for inhibitors and biosensors (Daye et al. 2010; Lamson et al. 2021).

The Shh affinity for heparin (Bumcrot et al. 1995) was initially attributed to the presence of a binding domain at the *N*-terminus (Rubin et al. 2002), more specifically a Cardin–Weintraub (CW) domain (Cardin and Weintraub 1989). Later research revealed a secondary heparin binding site involving basic residues in a globular domain (Farshi et al. 2011) and suggested the binding mode could be different in the monomeric and oligomeric forms of Shh. Structural elucidation of the murine Shh:heparin complex revealed that heparin could bind to a secondary binding site composed of residues Lys^{88}, Arg^{124}, Arg^{154}, Arg^{156}, although with lower affinity compared to the canonical CW residues Lys^{32}, Arg^{33}, Arg^{34}, Lys^{37}, Lys^{38} (Whalen et al. 2013).

This chapter focuses on the investigation of the Shh:heparin interaction with Epitopsy and MD. This research extends earlier work on the Shh:heparan sulfate interaction (Wilms 2013), which was instrumental to the genesis of Epitopsy. The following work also compares energy grids (EGs) obtained by Epitopsy against crystallographic data to assess the predictive power of the method.

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING
Cite This: J. Chem. Inf. Model. 2018, 58, 315−327
Article
pubs.acs.org/jcim

Locating Large, Flexible Ligands on Proteins

Jean-Noël Grad,[†] Alba Gigante,[‡] Christoph Wilms,[†] Jan Nikolaj Dybowski,[†] Ludwig Ohl,[†] Christian Ottmann,[§] Carsten Schmuck,[‡] and Daniel Hoffmann*[,†]

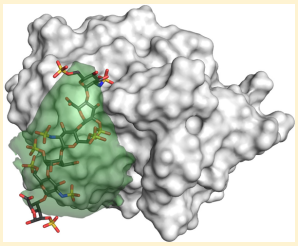
[†]Bioinformatics and Computational Biophysics, Faculty of Biology, University of Duisburg-Essen, Universitätstraße 7, 45141 Essen, Germany

[‡]Institute of Organic Chemistry, University of Duisburg-Essen, Universitätstraße 7, 45141 Essen, Germany

[§]Laboratory of Chemical Biology, Department of Biomedical Engineering, and Institute for Complex Molecular Systems, Eindhoven University of Technology, Den Dolech 2, 5612 AZ Eindhoven, The Netherlands

Supporting Information

ABSTRACT: Many biologically important ligands of proteins are large, flexible, and in many cases charged molecules that bind to extended regions on the protein surface. It is infeasible or expensive to locate such ligands on proteins with standard methods such as docking or molecular dynamics (MD) simulation. The alternative approach proposed here is scanning of a spatial and angular grid around the protein with smaller fragments of the large ligand. Energy values for complete grids can be computed efficiently with a well-known fast Fourier transform-accelerated algorithm and a physically meaningful interaction model. We show that the approach can readily incorporate flexibility of the protein and ligand. The energy grids (EGs) resulting from the ligand fragment scans can be transformed into probability distributions and then directly compared to probability distributions estimated from MD simulations and experimental structural data. We test the approach on a diverse set of complexes between proteins and large, flexible ligands, including a complex of sonic hedgehog protein and heparin, three heparin sulfate substrates or nonsubstrates of an epimerase, a multibranched supramolecular ligand that stabilizes a protein−peptide complex, a flexible zwitterionic ligand that binds to a surface basin of a Kringle domain, and binding of ATP to a flexible site of an ion channel. In all cases, the EG approach gives results that are in good agreement with experimental data or MD simulations.



INTRODUCTION

The prediction of binding poses of small molecules with a mixture of polar and hydrophobic groups that bind with high affinity in protein pockets has been one of the dominating problems in biomolecular modeling, and the successes in this endeavor have had a major impact in the life sciences and drug design. However, many biologically important interactions are almost the exact opposite of this scenario: large, flexible ligands bind to protein surfaces, their binding is often transient, and charge−charge interactions are essential. Examples are interactions between secreted proteins and the extracellular matrix of glycosaminoglycans,[1,2] interactions of virus proteins with host receptors in viral cell entry,[3] and interactions of T-cell receptors with MHC I−peptide complexes.[4] Another interesting case is the design of novel supramolecular ligands that bind protein surfaces with many low-affinity interactions but overall high avidity.[5]

How can we model and predict complexes of proteins with such large, flexible ligands that are often charged or zwitterionic? Sometimes it is possible to predict binding modes of large, flexible ligands by docking suitable fragments using methods developed for small-molecule docking.[6] This is less promising if binding occurs not in typical small-molecule binding pockets but rather at the protein surface, often involving charged residues with long, flexible side chains, as for instance in the case of protein−glycosaminoglycan binding. In these cases, interactions could be characterized by molecular dynamics (MD) simulation or related sampling methods,[7] though the necessary computational effort can be excessive.

Thus, current computational methods are either not suitable for the treatment of large, flexible ligands or are very expensive to apply. Unfortunately, the same also applies to high-resolution experimental characterization by X-ray crystallography or NMR spectroscopy. This makes it all the more important to develop reliable and efficient computational methods that can, e.g., be used to predict protein residues or regions that are crucial for the interaction with the ligand. Such predictions can then be tested, for instance, by measuring affinity changes after site-directed mutagenesis.

A promising avenue toward such efficient computational methods for this ligand class involves approaches that evaluate energies for ligand positions on a three-dimensional (3D) grid around the target protein. Although they have mainly been used for docking,[8] i.e., for locating optimal ligand positions and poses, in principle they allow for a characterization of the complete target protein surface region with respect to ligand binding energetics. A great advantage of grid-based approaches

Received: July 9, 2017
Published: December 21, 2017

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018, 58, 315−327

Journal of Chemical Information and Modeling Article

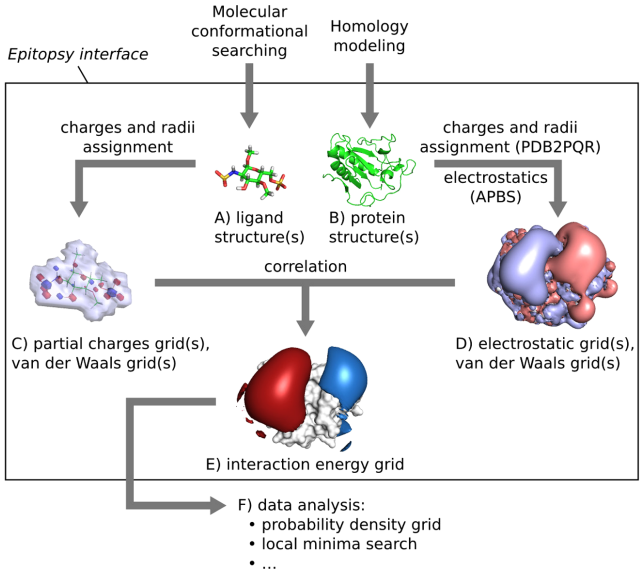


Figure 1. Workflow for the computation of (interaction) energy grids (EGs) with Epitopsy. Representative conformations of ligand fragments and the target protein are required as input (A, B). Charges and radii of both interaction partners are assigned (C, D). The EG for a ligand−protein pair as a "correlation" with FFTs is computed as described in text (E). EGs can then be analyzed in various ways outside Epitopsy (F).

is that the protein−ligand interaction energies on the grid can be evaluated efficiently by exploiting discrete fast Fourier transforms (FFTs), and this insight has led to the development of fast methods for protein−protein docking and for docking of small molecules to proteins.[9−13] However, these implementations of FFT-based docking have not been intended or used for locating large, highly flexible ligands on proteins.

Since we are mostly interested in interactions of proteins with charged ligands, another candidate method for characterizing the interaction energetics around the target protein is the solution of the Poisson−Boltzmann equation, typically also with efficient grid-based methods.[14,15]

In the work presented here, we assess the suitability of fast grid-based methods for predicting binding regions of large, flexible, and usually charged ligands on protein surfaces. These ligands not only defy docking methods but also force us to abandon the notion of the well-defined binding pose because their size and flexibility as well as the fact that they bind to extended protein surface regions will make binding more fuzzy.

One way to account for this uncertainty while still retaining a quantitative approach is to predict affinity distributions or probability densities for the ligand, or at least for those functional groups that likely mediate binding. The above-mentioned grid-based methods[9−12] are attractive because they could provide exactly this information in an efficient way. Generally, the approach proposed here assumes that we can infer the location of a large, flexible ligand from probability distributions of characteristic fragments and that these fragment probability distributions can be computed efficiently and sufficiently accurately by grid-based, FFT-accelerated scanning. We also demonstrate that flexibility of the target protein and of ligand fragments can be incorporated easily.

Many fragment-based approaches have been developed to identify binding poses or pockets of small drug-like molecules on proteins (see, e.g., refs 8 and 16−18), and some of these have even been applied to dock flexible peptides to well-defined binding pockets (see, e.g., ref 19), but to our knowledge, these methods are untried for our purpose of computing probability densities of our ligand class of interest on protein surfaces. We have therefore used MD as an established general-purpose reference method. An important advantage of comparison with MD simulations is that we can use a parametrization, especially charges and van der Waals radii, that is consistent between MD simulations and our "energy grid" (EG) computations, so that potential differences can be attributed to differences in conformational sampling and solvent model. Good agreement of EGs with X-ray structures and probability densities from MD simulations would be a first validation of EGs for our ligand class of interest. Since EGs can be computed at a small fraction of the computational cost of MD simulations, they could then be an interesting alternative approach.

To test the approach, we have evaluated EGs for five different test cases that cover several scenarios of practical interest: surface binding of heparin to sonic hedgehog protein, for which we compare several methods and experimental data; specific interactions of an epimerase with three different heparan sulfate substrates or nonsubstrates as an example of specificity of interaction; stabilization of a protein−peptide complex by an artificial multibranched supramolecular ligand as an example of a large nonpolymeric ligand; binding of a flexible zwitterionic ligand to a Kringle domain; and binding of ATP to a flexible receptor.

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315−327

Journal of Chemical Information and Modeling Article

METHODS

Workflow. Figure 1 shows the overall workflow of our grid-based analysis. Each of the steps will be described in the following sections. To make our results reproducible, we provide our experimental Epitopsy software as free open-source code at https://github.com/BioinformaticsBiophysicsUDE/Epitopsy.

To test the ability of the approach to predict binding regions of charged, flexible ligands on proteins, we have compared for several molecular systems electrostatic potentials (ESPs), EGs, probability densities, and crystal structures. For two of the systems, we have estimated probability densities of ligands by series of molecular dynamics simulations.

Molecular Dynamics Simulations. MD simulations were used first to determine representative ligand conformations as input to Epitopsy (Figure 1A) and second as a reference method to estimate probability densities of ligand fragments around the target protein that can then be compared with the corresponding probabilities computed from (interaction) energy grids generated by Epitopsy.

MD simulations were run with GROMACS 4.6.7[20] using the Amber ff99SB force field[21] for proteins and GlyCAM 06h-2[22] for saccharides. Phosphorylated serine parameters were obtained from the literature.[23] Nonstandard amino acids in ligand QQJ-096 (succinic acid, phenyl trihydrazine, *N*-acetyllysine, and GCP) were parametrized for the ff99SB force field according to the procedure described in the original ff94 article.[24] Atomic charges were derived from electrostatic potential maps calculated at the HF/6-31G* level of theory in Gaussian 09, revision A.02,[25] and fitted to the residues using the restrained electrostatic potential (RESP) method.[26,27] Force constant parameters were obtained by chemical analogy with readily available parameters in ff94.[28] Topology files were created with the pdb2gmx module of GROMACS for the protein and with the TLEaP module of Amber version 12.21[29] with the AmberTools suite version 13.22 for the ligands. Amber topologies were converted to GROMACS topologies by ACPYPE.[30]

Proteins and ligands were solvated in a dodecahedral box of SPC/E water molecules[31] with a minimum separation of 10 Å between the protein and the box boundaries. The system was neutralized by addition of Na^+ and Cl^- ions to a final ionic strength of 0.15 mol/L. The system was energy-minimized by steepest descent to a total force of 1800 kJ mol^{-1} nm^{-1} and equilibrated for 5 ns in the *NVT* ensemble with restrained heavy atoms followed by 5 ns in the *NPT* ensemble without restraints. Production simulations were run in the *NPT* ensemble for 250−500 ns. The temperature was stabilized at 300 K in the *NVT* and *NPT* ensembles using the V-rescale thermostat,[32] while the pressure was stabilized at 1 atm in the *NPT* ensemble using the Berendsen barostat (equilibration) or Parinello−Rahman barostat (data production).[33] Simulations were carried out on a GPU (GeForce 970 and GeForce 1070, CUDA 6.5) using a time step of 2 fs, the Verlet scheme[34] with a 10 Å cutoff for neighbor search, the particle mesh Ewald method[35] for electrostatic calculations, and the LINCS algorithm[36] for bond constraints.

Representative structures were extracted from trajectories on the basis of mutual RMSDs, using the g_rms tool in GROMACS to produce 2D RMSD plots, the PAM (partition around medoids[37]) tool from R package *cluster*, version 2.0.6, in R version 3.3.1[38] to find clusters, and the cluster.stats function from R package *fpc*, version 2.1.10, to validate the clustering on the basis of silhouette coefficients.[39]

When high flexibility in the ligand prevented the extraction of representative structures, the ligand trajectory was projected on a grid to produce a probability distribution of the ligand around the protein. To this end, the simulation box was discretized, and we counted for each grid point the number of MD frames where it was within the van der Waals radius of a ligand atom. The resulting count was divided either by the total number of frames in the trajectory to yield a grid point sampling frequency or by the sum of the grid point frequencies to yield a (ligand) probability density. The latter was used to compute cumulative density plots and to draw highest-density regions (HDRs)[40] in PyMOL. In comparisons of electrostatic, energy, and probability density grids, all of the compared grids were laid out with the same resolution, dimensions, and offset. EGs, HDRs, and molecules were visualized with PyMOL version 1.8[41] compiled from sources.

Protein Structures. Crystal structures were refined in MODELLER 9.17[42,43] to restore missing residues if necessary (Supporting Table S2). Candidate structures were required to minimize the DOPE and molpdf score. In the case of a tie, the refined model with lowest root-mean-square deviation (RMSD) with respect to the template was selected.

Assignment of Charges and Radii. For the computation of EGs, charges and radii have to be assigned to the ligand (Figure 1C) and protein (Figure 1D). Charges and atomic radii were added on the proteins with PDB2PQR version 2.0.0[44,45] at neutral pH and 298 K using the Amber force field option. Ligand charges were obtained from molecular mechanics (MM) force fields if available, or from one of these methods as specified in the text: electrostatic potential fit using the Merz−Singh−Kollman scheme[46,47] in Gaussian 09, revision A02,25 at the HF/6-31G** level of theory; or the Gasteiger−Marsili method[48,49] from OpenBabel[50] version 2.3.2.

Electrostatics. For the target protein, the electrostatic field was computed by solving the nonlinear Poisson−Boltzmann equation with APBS version 1.4.1[15] at 310 K with an ionic concentration of 0.15 mol/L and relative dielectric permittivities $\varepsilon_r^{\text{protein}} = 2$ and $\varepsilon_r^{\text{water}} = 79$.

Energy Grid Computation. The central part of the workflow is the computation of the energy grid for a ligand (or ligand fragment) and target protein (Figure 1E) by scanning the protein environment with the ligand or fragment. As we are mainly interested in charged ligands, the energy model currently only considers electrostatic interactions between the ligand and protein for non-overlapping relative positions and poses. EGs were calculated using the EnergyGrid tool of Epitopsy 1.0.[51] The following subsections describe how the energy is evaluated.

Shape Complementarity. The atomic description of a protein—obtained either from experimentally solved structures or from homology modeling—is mapped to a grid of dimensions (N_1, N_2, N_3) with resolution (m_1, m_2, m_3), usually in the range 0.5−1.0 Å. The default value in this work was 0.8 Å. Discretization proceeds by assigning a nonzero value to grid points within the van der Waals radii defined by PDB2PQR for protein and ligand atoms. These discretized geometries are labeled $f^{\text{vdw}}_{P_{l,m,n}}$ for the protein and $f^{\text{vdw}}_{L_{l,m,n}}$ for the ligand:

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315−327

$$\mathbf{f}^{\mathrm{vdw}}_{\mathrm{P}_{l,m,n}} = \begin{cases} \delta & \text{protein} \\ +1 & \text{surface layer} \\ 0 & \text{water} \end{cases}$$

$$\mathbf{f}^{\mathrm{vdw}}_{\mathrm{L}_{l,m,n}} = \begin{cases} +1 & \text{ligand} \\ 0 & \text{water} \end{cases} \quad (1)$$

The surface layer is the ensemble of solvent grid points in direct contact with the protein. The correlation is positive whenever the ligand is in contact with the protein surface (i.e., occupying the surface layer), negative when the ligand overlaps the protein, and zero otherwise. Ligand poses with negative shape correlation are discarded. Flexibility is introduced by the use of coefficients with opposite sign: an overlapping pose with n overlapping grid points is rejected unless a minimum of $|\delta \cdot n|$ grid points are in surface contact. We used mainly $\delta = -15$ as given by ref 10, but we point out in the Discussion and Supporting Figure S4 that it can be useful to vary δ.

The shape correlation $\mathbf{f}^{\mathrm{vdw}}_{\mathrm{C}_{\alpha,\beta,\gamma}}$ can be defined as the direct product of the two matrices $\mathbf{f}^{\mathrm{vdw}}_{\mathrm{P}_{l,m,n}}$ and $\mathbf{f}^{\mathrm{vdw}}_{\mathrm{L}_{l,m,n}}$ for any shift vector (α, β, γ):

$$\mathbf{f}^{\mathrm{vdw}}_{\mathrm{C}_{\alpha,\beta,\gamma}} = \sum_{l}^{N_1} \sum_{m}^{N_2} \sum_{n}^{N_3} \mathbf{f}^{\mathrm{vdw}}_{\mathrm{P}_{l,m,n}} \mathbf{f}^{\mathrm{vdw}}_{\mathrm{L}_{l+\alpha,m+\beta,n+\gamma}} \quad (2)$$

This calculation has an asymptotic time complexity of $O(n^6)$, making it impractical for solving numerically large systems. The fast Fourier transform $\mathcal{F}$ (or $\mathcal{F}^{-1}$ for the reverse operation) was successfully introduced by Gabb et al.[10] in this context, resulting in a time complexity of $O(n^3 \ln(n^3))$:

$$\begin{aligned} \mathbf{F}^{\mathrm{vdw}}_{\mathrm{P}} &= \mathcal{F}\{\mathbf{f}^{\mathrm{vdw}}_{\mathrm{P}_{l,m,n}}\} \\ \mathbf{F}^{\mathrm{vdw}}_{\mathrm{L}} &= \mathcal{F}\{\mathbf{f}^{\mathrm{vdw}}_{\mathrm{L}_{l,m,n}}\} \\ \mathbf{F}^{\mathrm{vdw}}_{\mathrm{C}} &= \overline{\mathbf{F}^{\mathrm{vdw}}_{\mathrm{P}}} \mathbf{F}^{\mathrm{vdw}}_{\mathrm{L}} \\ \mathbf{f}^{\mathrm{vdw}}_{\mathrm{C}_{\alpha,\beta,\gamma}} &= \mathcal{F}^{-1}\{\mathbf{F}^{\mathrm{vdw}}_{\mathrm{C}}\} \end{aligned} \quad (3)$$

where the uppercase letter $\mathbf{F}$ represents the decomposed signal $\mathbf{f}$ and $\overline{\mathbf{F}}$ is the complex conjugate of $\mathbf{F}$.

Electrostatic Energy. The electrostatic potential (ESP) $\mathbf{\Phi}_{l,m,n}$ of the protein in ionic aqueous solution obtained by solving the nonlinear Poisson−Boltzmann equation (see above) is stored in a matrix $\mathbf{\Phi}_{\mathrm{P}_{l,m,n}}$, with the protein interior and surface set to a potential of zero, and the matrix $\mathbf{q}_{\mathrm{L}_{l,m,n}}$ contains the ligand partial charges:

$$\mathbf{\Phi}_{\mathrm{P}_{l,m,n}} = \begin{cases} 0 & \text{protein} \\ \mathbf{\Phi}_{l,m,n} & \text{water} \end{cases}$$

$$\mathbf{q}_{\mathrm{L}_{l,m,n}} = \begin{cases} \mathbf{q}_{l,m,n} & \text{ligand} \\ 0 & \text{water} \end{cases} \quad (4)$$

The electrostatic interaction between the ligand partial charges and the protein electrostatic potential is used to compute the energy correlation matrix $\mathbf{f}^{\mathrm{elec}}_{\mathrm{C}_{\alpha,\beta,\gamma}}$:

$$\mathbf{f}^{\mathrm{elec}}_{\mathrm{C}_{\alpha,\beta,\gamma}} = \sum_{l}^{N_1} \sum_{m}^{N_2} \sum_{n}^{N_3} \mathbf{\Phi}_{\mathrm{P}_{l,m,n}} \mathbf{q}_{\mathrm{L}_{l+\alpha,m+\beta,n+\gamma}} \quad (5)$$

The same FFT optimization as described in eq 3 is used to speed up the correlation here:

$$\begin{aligned} \mathbf{F}^{\mathrm{elec}}_{\mathrm{P}} &= \mathcal{F}\{\mathbf{\Phi}_{\mathrm{P}_{l,m,n}}\} \\ \mathbf{F}^{\mathrm{elec}}_{\mathrm{L}} &= \mathcal{F}\{\mathbf{q}_{\mathrm{L}_{l,m,n}}\} \\ \mathbf{F}^{\mathrm{elec}}_{\mathrm{C}} &= \overline{\mathbf{F}^{\mathrm{elec}}_{\mathrm{P}}} \mathbf{F}^{\mathrm{elec}}_{\mathrm{L}} \\ \mathbf{f}^{\mathrm{elec}}_{\mathrm{C}_{\alpha,\beta,\gamma}} &= \mathcal{F}^{-1}\{\mathbf{F}^{\mathrm{elec}}_{\mathrm{C}}\} \end{aligned} \quad (6)$$

Correlation Matrix. The energy correlation matrix $\mathbf{f}^{\mathrm{elec}}_{\mathrm{C}_{\alpha,\beta,\gamma}}$ (eq 5) is used to calculate $\mathbf{\Delta E}^{\mathrm{elec}}_{\mathrm{bind}}(\alpha, \beta, \gamma)$, the electrostatic contribution to the binding affinity for any shift vector (α, β, γ) where the molecular probe does not overlap with the protein, i.e., for $\mathbf{f}^{\mathrm{vdw}}_{\mathrm{C}_{\alpha,\beta,\gamma}} \geq 0$:

$$\mathbf{\Delta E}^{\mathrm{elec}}_{\mathrm{bind}}(\alpha, \beta, \gamma) = \mathbf{f}^{\mathrm{elec}}_{\mathrm{C}_{\alpha,\beta,\gamma}} H[\mathbf{f}^{\mathrm{vdw}}_{\mathrm{C}_{\alpha,\beta,\gamma}}] \quad (7)$$

where $H[x]$ is the Heaviside step operator, given by

$$H[x] = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (8)$$

Angular Sampling and Energy Grid Values. The correlation matrices are evaluated for many orientations $\boldsymbol{\omega} \in \mathbf{\Omega}$ of the ligand, where $\mathbf{\Omega}$ is a set of rotation angles (ϕ, θ, ψ) with (ϕ, θ)-tuples uniformly distributed on a sphere using a Fibonacci generative spiral[52,53] and ψ an optional spin around each (ϕ, θ) tuple, sampled uniformly in $[0, 2\pi]$ with a step of $2\pi/n$ ($n = 1$ for no spin or any nonzero even number for a spin). The binding free energy $\mathbf{\Delta G}^{\mathrm{elec}}_{\mathrm{bind}}(\alpha, \beta, \gamma)$ is computed from $|\mathbf{\Omega}|$ correlations:

$$\mathbf{\Delta G}^{\mathrm{elec}}_{\mathrm{bind}}(\alpha, \beta, \gamma, \mathbf{\Omega}) = -k_{\mathrm{B}}T \ln\left(\frac{\sum_{\boldsymbol{\omega} \in \mathbf{\Omega}} \exp(-\mathbf{f}^{\mathrm{elec}}_{\mathrm{C}^{\boldsymbol{\omega}}_{\alpha,\beta,\gamma}} H[\mathbf{f}^{\mathrm{vdw}}_{\mathrm{C}^{\boldsymbol{\omega}}_{\alpha,\beta,\gamma}}]/k_{\mathrm{B}}T)}{|\mathbf{\Omega}|} \right) \quad (9)$$

The division by $|\mathbf{\Omega}|$ accounts for the purely entropic free energy of the reference state, i.e., the ligand immersed in pure solvent, where it takes $|\mathbf{\Omega}|$ orientations of zero enthalpy. We used mainly $|\mathbf{\Omega}| = 150$ without spin, as this value provides a reasonable trade-off between accuracy and calculation time; we show in Supporting Figure S3 the effect of increasing $|\mathbf{\Omega}|$.

The number of available ligand rotations at every grid point is

$$\mathbf{\Omega}^{\mathrm{available}}(\alpha, \beta, \gamma, \mathbf{\Omega}) = \sum_{\boldsymbol{\omega} \in \mathbf{\Omega}} H[\mathbf{f}^{\mathrm{vdw}}_{\mathrm{C}^{\boldsymbol{\omega}}_{\alpha,\beta,\gamma}}] \quad (10)$$

The ligand excluded volume (LEV) corresponds to the set of grid points (α, β, γ) where no rotational state with finite energy is available to the ligand ($\mathbf{\Omega}^{\mathrm{available}} = 0$), or the set of all grid points where the function **LEV** is 1:

$$\mathbf{LEV}(\alpha, \beta, \gamma, \mathbf{\Omega}) = 1 - H[\mathbf{\Omega}^{\mathrm{available}} - 1]$$

Energy Grids for Multiple Conformers. When several conformers $\mathbf{P}$ of the protein and $\mathbf{L}$ of the ligand are provided, with respective internal energies U_i for $\mathbf{P}$ and U_j for $\mathbf{L}$, the binding free energy is

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018, 58, 315−327

Journal of Chemical Information and Modeling Article

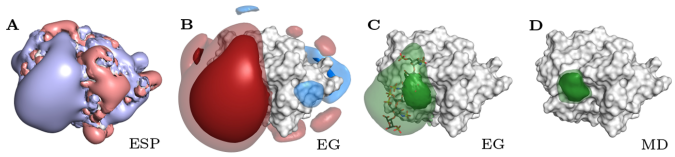


Figure 2. Sonic hedgehog protein (Shh) with heparin ligand. (A) ESP isosurfaces of Shh at $+1k_BT/|e|$ (blue) and $-1k_BT/|e|$ (red). (B) EG isosurfaces at $\pm 1k_BT$ (translucent blue/red) and $\pm 2k_BT$ (solid blue/red), merging across two population-weighted conformations of a heparin disaccharide (clustering details and glycosidic angles are provided in Supporting Table S1) according to eq 11 using $|\mathbf{\Omega}| = 600$ and $n = 4$. (C) EG-based probability densities of heparin disaccharide drawn around 20% HDR (solid green) and 30% HDR (translucent green). The 20% HDR forms a hull around the crystallographic position of the heparin tetrasaccharide from ref S6 (PDB entry 4c4n). (D) MD-based probability densities of heparin disaccharide drawn at 20% and 30% HDR (solid and translucent green) from a 3.5 μs multitrajectory MD simulation. MD starting conditions and trajectories are detailed in Supporting Figure S2.

$$\Delta \mathbf{G}^{\mathrm{elec}}_{\mathrm{bind}}(\alpha, \beta, \gamma, \mathbf{P}, \mathbf{L}, \mathbf{\Omega}) = -k_BT \ln\left(\frac{\sum_{i\in\mathbf{P}} e^{-U_i/k_BT} \sum_{j\in\mathbf{L}} e^{-U_j/k_BT} \sum_{\omega\in\mathbf{\Omega}} \exp(-\mathbf{f}^{\mathrm{elec}}_{C^{ij\omega}_{\alpha,\beta,\gamma}} H[\mathbf{f}^{\mathrm{vdw}}_{C^{ij\omega}_{\alpha,\beta,\gamma}}]/k_BT)}{|\mathbf{\Omega}| \cdot \sum_{i\in\mathbf{P}} e^{-U_i/k_BT} \sum_{j\in\mathbf{L}} e^{-U_j/k_BT}}\right) \quad (11)$$

The number of available orientations is

$$\mathbf{\Omega}^{\mathrm{available}}(\alpha, \beta, \gamma, \mathbf{P}, \mathbf{L}, \mathbf{\Omega}) = \sum_{i\in\mathbf{P}} \sum_{j\in\mathbf{L}} \sum_{\omega\in\mathbf{\Omega}} H[\mathbf{f}^{\mathrm{vdw}}_{C^{ij\omega}_{\alpha,\beta,\gamma}}] \quad (12)$$

Conversion to Probability Densities. EGs may be transformed into probability density functions (PDFs) with Boltzmann factors **K** for positions outside the LEV:

$$\mathbf{K}(\alpha, \beta, \gamma) = [1 - \mathbf{LEV}(\alpha, \beta, \gamma, \mathbf{\Omega})] \exp\left(\frac{-\Delta \mathbf{G}^{\mathrm{elec}}_{\mathrm{bind}}(\alpha, \beta, \gamma, \mathbf{\Omega})}{k_BT}\right)$$

$$\mathbf{PDF}(\alpha, \beta, \gamma) = \frac{\mathbf{K}(\alpha, \beta, \gamma)}{\sum \mathbf{K}(\alpha, \beta, \gamma)} \quad (13)$$

Quantitative Comparison of EGs with Experimental Ligand Positions and MD. To obtain a simple measure of agreement between ligand positions in crystal structures and EGs, we used the following method. The EG was partitioned into clusters corresponding to local energy minima with a series of several hundred simulated annealing (SA) runs, in each letting a probe walk on the EG to an energy minimum.[54,55] The minima were postprocessed with a robust silhouette-validated PAM clustering.[37,39] With this procedure, we effectively denoised the EG and ended up with a small number of compact clusters—always less than 20 clusters in the cases studied here. We further assigned to the cluster the energy value of its medoid. We then took the distance between the cluster medoids and the geometric center of the ligand in the crystal structure as a measure of agreement between the EGs and the crystal structure: the smaller the distance $d_{\mathrm{X-EG}}$ between the EG cluster medoid of lowest energy and the crystal structure, the better the agreement between the EG and the crystal structure.

For comparison between MD results and EGs, we first computed the EGs and determined clusters using the SA/clustering approach described above. Then we used the same EG grid geometry to discretize the MD trajectories: For each ligand pose sampled in MD, a counter at each grid point within the ligand volume was incremented by 1. In this way, the complete set of trajectories was mapped onto counts of grid point hits. These were then transformed into probabilities by normalization, which were log-transformed into energies. The SA/clustering approach was applied to this MD-derived grid in the same way as described above. Finally, we evaluated the distances $d_{\mathrm{MD-EG}}$ between the MD and EG medoids of lowest energy or the distances between neighboring minima in either grid.

RESULTS

Sonic Hedgehog and Heparin. The complex of sonic hedgehog protein (Shh) and a heparin ligand is prototypical of our systems of interest: a large, flexible, and highly charged ligand binds to the surface of a protein. The general assumption underlying our computational assessment of the heparin location is that we can infer the location of large, flexible ligands from the probability densities of characteristic fragments. Of course, this assumption has to be tested, and it will break down under certain conditions, as we outline in the Discussion. To test the approach, we therefore compiled for this system a comprehensive data set consisting of the crystal structure of the Shh–heparin tetrasaccharide complex from ref S6, the electrostatic potential (ESP) of Shh, the energy grid (EG) and probability densities for disaccharide heparin fragment scans of Shh with Epitopsy, and seven 500 ns MD simulations of Shh with disaccharide fragments (placed initially at random positions on the Shh surface; Supporting Figure S2). The MD trajectories were sufficiently long to observe ligand binding and unbinding events (Supporting Figure S9).

Overall, the four sources of data give a consistent picture (Figure 2): The ESP has its largest high-potential region around Arg156, and this is where the EG has its largest low-energy blob and where the heparin tetrasaccharide is located in the crystal structure. This is also the area of the highest heparin disaccharide probability density, as estimated from MD trajectories, with good agreement between the MD and EG global minima ($d_{\mathrm{MD-EG}} = 3.3$ Å).

Assuming a Boltzmann distribution, the EG and ESP values can be transformed into probability densities for ligand occupancy (eq 13). These probability densities can then be compared directly with the probability densities estimated by MD sampling, either visually (e.g., with 3D isosurfaces; Figure 2) or quantitatively (Figure 3). For the latter, we evaluated the

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315–327

Journal of Chemical Information and Modeling Article

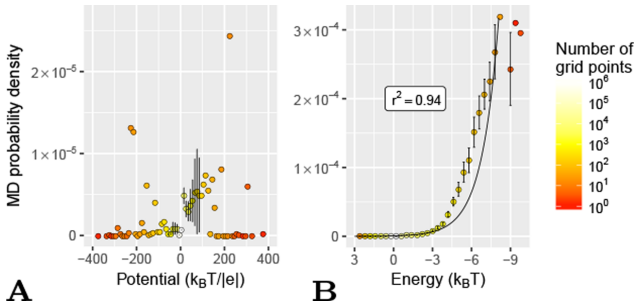


Figure 3. Probability density of heparin disaccharide occupancy computed from MD simulations vs (A) electrostatic potential and (B) energy grid at the same grid positions. Each point in each plot stands for all grid points with a certain value of (A) potential or (B) energy as given by its position along the horizontal axis. The numbers of grid points with the respective ESP or EG values are shown as colors. Both horizontal axes go from repulsive to attractive, and in both panels the vertical axes give the probability estimated by MD sampling averaged over the grid points with a given ESP or EG value. The error bars mark the 99% confidence intervals assuming normally distributed probabilities. In (A), 13 outliers in the ESP grid with energies ranging from $400k_BT/|e|$ to $800k_BT/|e|$ lie outside the plotting range. Supporting Figure S11 shows the distribution in a 2D histogram.

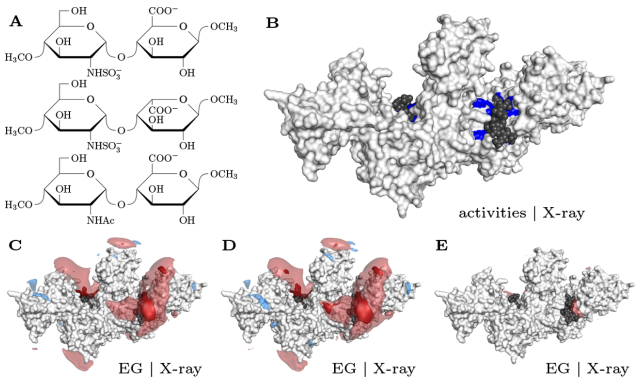


Figure 4. Interaction of D-glucuronyl C_5-epimerase with heparan sulfate. (A) Disaccharides used to compute EGs around epimerase: (top) substrate $CH_3O-GlcNS-GlcA-OCH_3$; (middle) product and in vitro substrate $CH_3O-GlcNS-IdoA-OCH_3$; (bottom) nonsubstrate $CH_3O-GlcNAc-GlcA-OCH_3$. (B) Crystal structure of epimerase in complex with heparin hexamer (PDB entry 4pxq[58]). The two heparin fragments (black) bind at the two active sites of the C_2-symmetric enzyme dimer. Amino acids critical for reaction (Ala mutations lead to enzyme activity loss of >60% compared with the wild type[58]) are marked in blue. (C–E) EGs of the substrate (C), product (D), and nonsubstrate (E) scanned using the apo protein (PDB entry 4pw2[58]) with isosurfaces drawn at $\pm 1k_BT$ (translucent blue/red) and $-2k_BT$ (solid red) and crystallographic heparin (black space-filling). The isosurfaces in (C–E) were robust against changes in dihedral angles of the HS dimer used for scanning. Glycosidic angles are provided in Supporting Table S1.

frequency with which the heparin disaccharide visited each EG/ESP grid cube in the concatenated MD trajectories as described in Methods. The highest probability densities from EG and MD are located in the same area around Arg156 (Figure 2C,D), where the EG shows its by far largest low-energy blob (Figure 2B). However, the EG probability density maximum there has a much larger spatial spread than the MD probability density. Interestingly, the region of 20% highest probability density as computed from heparin disaccharide EGs forms an envelope around the crystal position of the heparin tetrasaccharide, following the crystal ligand in shape and size (Figure 2C). The center of the heparin hexamer in the crystal structure is closest to the lowest-energy cluster medoid of the EG from the scan with the heparin disaccharide. The distance is d_{X-EG} = 6.1 Å (see the last paragraph of Methods and Supporting Figure S10A), i.e., less than two monosaccharide units. It should be noted that the distance of the medoid to the closest disaccharide subunit of the crystal ligand is 4.1 Å, i.e., less than one monosaccharide away.

In another quantitative comparison between the EG and MD probability densities (Figure 3B), we see that the MD probability density roughly follows an exponential of the EG values, as expected for a Boltzmann distribution (coefficient of determination r^2 = 0.94). The deviation between the actual distribution and an exponential could be a result of unequilibrated MD sampling or EG model deficiencies.

The obvious similarity of ESP and EG (Figure 2A,B) suggests that ESP should have a similarly good association with MD. However, this is not the case (Figure 3A). If we transform ESPs into probability densities for a charged ligand, the probability density is almost completely concentrated at a single grid point close to the two-calcium center of Shh, 2 nm away

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315–327

from Arg156. While this point is certainly very attractive for the heparin disaccharide if we only consider Coulomb interactions, it is sterically not accessible and therefore neither visible in the EG nor sampled by MD. Figure 3A suggests that the same is true for many points of high ESP that are barely explored in MD simulations or evaluated in the EG. On the other hand, excluding points that may be sterically difficult to access from the ESP evaluation also makes the information obtained from ESP alone much less specific (see Supporting Figure S8).

C_5-epimerase and Polyanionic Heparan Sulfate Substrates and Nonsubstrates. D-Glucuronyl C_5-epimerase modifies heparan sulfate (HS), i.e., long, negatively charged, and highly flexible carbohydrate chains. The epimerase has a varied surface topography with deep clefts. The HS chains have to be threaded through a narrow, partially buried active site, which makes the epimerase−HS complex a harder test case than the Shh−heparin complex of the previous section, where heparin bound preferentially to a well-accessible surface patch on Shh. The more specific, conformation-dependent chemical function of the epimerase suggests a more accurate positioning of the HS substrate chains on epimerase than the superficial attachment of heparin to Shh. The hypothesis of a more accurate positioning is consistent with the observed substrate length dependence of the reaction: the enzymatic activity decreased by 90% on a digested heparan sulfate fraction containing octasaccharides and smaller oligosaccharides.[57] Our question was therefore whether we would be able to trace an extended binding site in EGs that could accommodate such longer oligosaccharides. For validation, we compared the predicted binding sites with crystallographically determined binding sites of an inhibitor, a heparin hexamer (Protein Data Bank (PDB) entry 4pxq[58]).

We used three different HS dimer fragments (Figure 4A) to compute the EGs: $CH_3O-GlcNS-GlcA-OCH_3$ as model of the substrate, $CH_3O-GlcNS-IdoA-OCH_3$ as model of the product, and $CH_3O-GlcNAc-GlcA-OCH_3$ as a nonsubstrate.[57,59] It should be noted that in vitro the enzyme works both ways, i.e., the product is a substrate for the reverse reaction.[59] A parsimonious natural explanation of this finding is that the substrate and product use the same molecular binding site.

In fact, in the EGs with epimerase substrate and product we detected the same low-energy channel, centered around the active sites (Figure 4B−D). The region that binds most strongly in the EG matches the crystallographic positions of the heparin hexasaccharide and covers the amino acid residues most important for enzymatic activity.[58] For all three ligands, the crystal position is closest to the lowest-energy cluster medoid of the EG (d_{X-EG} = 2.6, 3.1, and 4.5 Å; also see Supporting Figure S10E,F). However, the low-energy region extends noticeably beyond the crystallographic location of the heparin hexasaccharide and could easily accommodate HS oligomers longer than octasaccharides (translucent red in Figure 4C,D). The shape of this low-energy region suggests a core binding site for HS chains reaching from the right flank of the narrow cleft with the active center down the crystallographic heparin binding site.

While the substrate and product are both doubly negatively charged, the nonsubstrate molecule (bottom of Figure 4A) carries only one negative charge. In the corresponding EG, the $-1k_BT$ region has shrunk drastically and now only covers the location of the crystallographic heparin hexasaccharide. Thus, although the nonsubstrate could be chemically epimerized in principle—it has the same GlcA amenable to epimerization—this particular epimerase enzyme offers no suitable binding site for a longer chain of this nonsubstrate type.

14-3-3 Protein and Polycationic Supramolecular Ligand. Recently we were able to demonstrate experimentally (Gigante et al., unpublished) that the binding of a supramolecular ligand, QQJ-096[60] (Figure 5B), stabilizes the interaction between the 14-3-3 protein and peptide fragments of c-Raf protein (we call this complex 14-3-3/c-Raf). The large QQJ-096 ligand has three flexible arms ("R" in Figure 5B), each of which ends in two positively charged groups, a lysine and a guanidinocarbonylpyrrole (GCP). The size and flexibility of the ligand makes it unsuitable for small-molecule docking, and it is unlikely that this ligand takes a single, well-defined binding pose. Nevertheless, its effect on the interaction of c-Raf and 14-

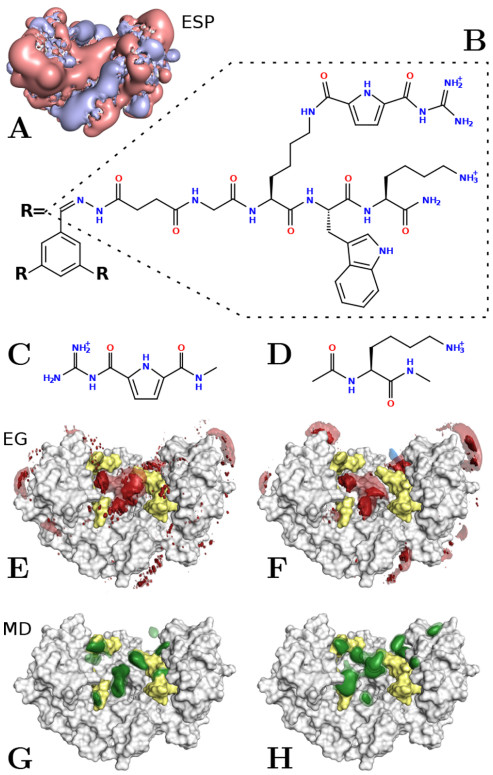


Figure 5. Interaction of 14-3-3/c-Raf complex with the supramolecular ligand QQJ-096. (A) Isosurfaces of electrostatic potential at $+1k_BT/|e|$ (blue) and $-1k_BT/|e|$ (red) of 14-3-3/c-Raf (PDB entry 4ihl[61]). (B) QQJ-096 ligand with only one of three arms ("R") shown. (C) GCP with capped ends. (D) Lys with capped ends. (E−H) 14-3-3/c-Raf protein complex (14-3-3 in white, c-Raf in yellow): (E, F) EGs for GCP and Lys, respectively, with isosurfaces drawn at $\pm1k_BT$ (translucent blue/red) and $-1.5k_BT$ (solid red); (G, H) 20% HDR (solid green) and 30% HDR (translucent green) for GCP and Lys, respectively, from a 1.5 μs multitrajectory MD simulation of the 14-3-3/c-Raf complex with QQJ-096 in aqueous solution.

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315−327

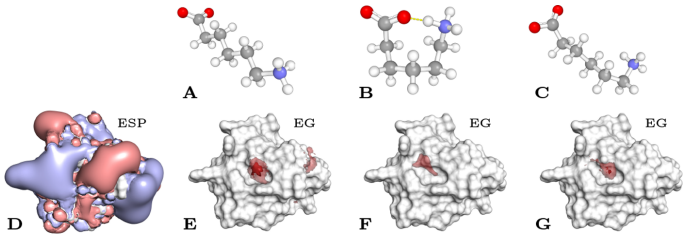


Figure 6. ε-Aminocaproic acid (EACA) and its complex with plasminogen Kringle domain 4 (KR4). (A) Experimental structure ("stretched") of EACA in the solid phase (CCDC entry 1102509[63]). (B) Energy-minimized structure of EACA ("turn", MMFF94 force field, steepest descent in vacuum). (C) Structure of EACA in complex with KR4 (PDB entry 2pk4). (D) Electrostatic potential isosurfaces around apo KR4 (PDB entry 2pk4) drawn at $+1k_BT/|e|$ (blue) and $-1k_BT/|e|$ (red). (E–G) EG isosurfaces drawn at $\pm1k_BT$ (translucent blue/red) and $\pm2k_BT$ (solid blue/red) of apo KR4 with EACA stretched (E), as turn (F), and merging across multiple conformations of the ligand and protein according to eq 11 (G). For the ligand Gasteiger–Marsili partial charges were used.

3-3 could be explained most easily by specific binding of QQJ-096 to 14-3-3/c-Raf.

The electrostatics of 14-3-3/c-Raf shows many regions of low electrostatic potential (red in Figure 5A) that could interact with the positive end groups of QQJ-096, i.e., GCP (Figure 5C) and Lys (Figure 5D). For a more ligand-specific assessment of binding, we computed two EGs with molecules corresponding to these two end groups. The EGs show roughly the same features for GCP (Figure 5E) and Lys (Figure 5F), with particularly high affinities in the center of the 14-3-3 cleft between the c-Raf peptides. For Lys there are additional high-affinity patches so that the c-Raf fragments are sandwiched between regions of high affinity for Lys. Both end groups have a few small high-affinity islands outside the central cleft of 14-3-3/c-Raf, with Lys having more of those islands than GCP. Overall, this result suggests that QQJ-096 could seal off the 14-3-3/c-Raf cleft and in this way inhibit dissociation of c-Raf fragments, in agreement with experimental results (Gigante et al., unpublished).

For comparison we simulated the 14-3-3/c-Raf/QQJ-096 system with MD. Because of the size of the molecular system and the low charge density on the ligand, a computational experiment analogous to the Shh–heparin experiment above seemed to be unfeasible, i.e., we did not expect to reach the MD steady state on the microsecond time scale with a ligand initially positioned at random in the solvent box. On the basis of the experimental evidence for QQJ-096-mediated stabilization of the 14-3-3/c-Raf complex and the assumption of a direct mode of interaction, the MD starting conditions could be narrowed down to the 14-3-3 cleft (Supporting Figure S1). On the basis of this reasoning, we carried out a pilot set of six 50 ns MD simulations of the 14-3-3/c-Raf dimer in aqueous solution with QQJ-096 initially positioned 10 Å above the c-Raf peptides (Supporting Figure S1A,C). In three simulations the ligand failed to interact with the protein. We then ran six 250 ns simulations with the ligand initially positioned 4–6 Å above the c-Raf peptides (Supporting Figure S1B,C) and observed a quick convergence to binding sites of QQJ-096 end groups in the 14-3-3 cleft close to those predicted by EGs computed with the end groups (Figure 5G,H and Supporting Table S3). Regions outside the cleft were barely explored. Thus, the MD simulations and EGs support the same mechanism for the experimentally observed stabilization of 14-3-3/c-Raf binding by QQJ-096, namely, that the supramolecular ligand QQJ-096 blocks the 14-3-3 cleft and in this way impedes escape of c-Raf.

QQJ-096 is a case where the comparison of MD and EGs revealed an interaction of the Lys and GCP end groups on the same ligand. While the distributions of GCP and Lys are similar, as can be seen in the EGs separately computed for each of the two fragments (Figure 5E,F), the distances between the MD-derived medoids and the closest EG-derived medoids (see the last section of Methods) are smaller for GCP than for Lys (Supporting Table S3). If we consider that GCP groups bind more strongly to negatively charged groups than Lys, this observation can be readily explained as displacement of Lys by GCP.

Kringle Domain and Flexible Zwitterionic Ligand. The Kringle domains of plasminogen attach to Lys residues on fibrin, a precondition for the decomposition of fibrin by plasminogen. A known alternative ligand of the Kringle domains is ε-aminocaproic acid (EACA), and a crystal structure of its complex with plasminogen Kringle domain 4 (KR4) has been determined (PDB entry 2pk4[62]). EACA is a highly flexible zwitterionic molecule (Figure 6A–C) that binds to a shallow basin in the KR4 surface. We used the EACA–KR4 complex as a test case for the application of EGs based on multiple ligand conformers for the identification of binding sites (eq 11). In an application scenario we would probably not know the actual conformer but instead would rely on plausible ligand conformers obtained from other experiments or modeling. Accordingly, our EACA input conformers were the stretched conformer (Figure 6A) observed in the solid phase of pure EACA[63] and an MM-energy-minimized turn geometry (Figure 6B) that is entropically and enthalpically more favorable for a free ligand. The conformer actually observed in the crystal complex (Figure 6C) is closer to the stretched geometry in the solid phase (Figure 6A), though with a bent amino end.

The zwitterionic nature of EACA suggests a binding site that bridges two regions of opposite electrostatic potential. However, this pattern is too unspecific since there are many regions that fall into this category (Figure 6D). The full correlation with shape and electrostatics information leads to the identification of the correct binding basin, which in fact bridges regions of opposite electrostatic potential (Figure 6E–G). For the stretched EACA conformer there are two binding sites: the one in the crystal structure and an alternative binding site with slightly lower affinity between Asp381 and Lys433 (Figure 6E). For the EACA turn conformer, the correct basin is clearly the region with highest affinity (Figure 6F). The EG averaged over both ligand conformers (eq 11, with the two

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315–327

ligand conformations weighted equally) also has the basin of the crystal structure as the clearly dominant binding site (Figure 6G). The crystal ligand is closest to the lowest-energy cluster medoid of the EG (d_{X-EG} = 1.7 Å; also see Supporting Figure S10D).

A question related to the multiconformer ligand treatment is the multiconformer receptor treatment, and eq 11 treats the ligand and receptor symmetrically in this respect. In fact, Figure 6G is based not only on two ligand conformations but also on three equally weighted KR4 receptor conformations, including the EACA–KR4 complex structure (2pk4), a KR4 complex with sulfate (PDB entry 1krn), and a KR4 complex with arginine (PDB entry 4duu). However, since the differences between the receptor structures are small (RMSDs of C_α atoms with respect to 2pk4: 0.29 Å for 1krn and 0.69 Å for 4duu), the results are very similar for the combined EG and for the three EGs based on single receptor structures: for all three apo KR4 structures, the respective EGs identified the same correct EACA binding basin as the best binding site.

Flexible ATP Binding Site of the P2X$_4$ Ion Channel. In some applications of EGs with flexible protein receptors, either the apo state or the holo state of the protein is known, but not both, leading to some uncertainty as to whether EGs for one state allow inferences for the other state. An example where this question can be tested is the ATP-gated P2X$_4$ receptor cation channel. This protein undergoes a significant conformational change upon binding of ATP, and crystal structures of the apo and holo states of P2X$_4$ (PDB entries 4dw0 and 4dw1, respectively[64]) show remarkable plasticity at the ATP binding site (Figure 7A). To evaluate the resilience of EGs toward the conformational change of the receptor protein, the apo and holo structures of P2X$_4$ were evaluated with ATP as the ligand. The two structures led to similar EGs (Figure 7C,D) that correctly identified the crystallographic ATP binding site (Figure 7B). Although P2X$_4$ is a transmembrane protein, it was completely solvated for simplicity. Correlation information at the transmembrane interface, which is 40 Å away from the ATP binding site, should therefore be ignored (this region contributes very little since it is nonpolar). The absolute minimum of the EG is located at the actual ATP binding site for both the apo and holo structures (Figure 7C,D). In the holo structure of the receptor, the probability density of ATP is sharply concentrated to only a few grid points at the crystallographic binding site (Figure 7F), while in the apo structure the highest probability density is more spread out around this point (Figure 7E). The positions of the three crystal ligands in both the apo and holo structures are closest to the lowest-energy cluster medoids of the EG (d_{X-EG} = 3.8–6.8 Å; also see Supporting Figure S10B,C).

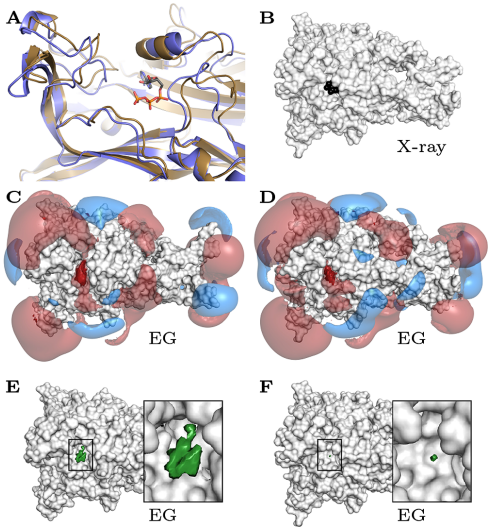


Figure 7. P2X$_4$ channel and its interaction with ATP. (A) Superposition of the ATP-free apo (blue, PDB entry 4dw0[64]) and ATP-bound holo (brown, ATP in sticks, PDB entry 4dw1) crystal structures of P2X$_4$. (B) Crystal structure of the P2X$_4$–ATP complex; the ligand (black spheres) binds at the interface between two chains of the C_3-symmetric trimer. (C, D) EG isosurfaces of ATP drawn at $\pm 1 k_B T$ (translucent blue/red) and $\pm 8 k_B T$ (solid blue/red) for the (C) apo and (D) holo structures of P2X$_4$. (E, F) EG-based probability densities of ATP drawn around 95% HDR, with zoomed-in panels showing only the binding site. The same was observed for the two other C_3-related binding sites.

DISCUSSION

The very nature of large, flexible ligands, such as glycosaminoglycan chains, large receptor loops, and novel supramolecular binders, makes it challenging to model their interactions with proteins. First, these ligands are too large and flexible for small-molecule docking. Even the notion of a well-defined binding pose, commonly used in small-molecule docking, has to be abandoned. Instead, we should restate the aim from finding the binding pose to computing a probability density for the ligand around the protein. Second, the vastness of their conformational spaces makes large, flexible ligands also difficult objects for the standard method MD simulation, as it will be prone to severe undersampling.[65,66] A third established candidate method is continuum electrostatics. It takes advantage of the charged nature of the ligands, but the electrostatic potential alone is difficult to interpret or insufficiently specific. The fact that electrostatics alone is not sufficient to locate large charged ligands like DNA or RNA has been noted before.[67,68] Our results corroborate this because we found only a weak correlation between the electrostatic potential, the probability density obtained from extensive MD simulations of the Shh/heparin system, and corresponding experimental data. In this case, we achieved good consistency of MD simulation and experimental structures with EGs that complement electrostatics with information about shape, charge distribution, and volume of the ligand or ligand fragments.

The good correlation of the EG approach with probability densities from extensive MD simulations at a small fraction of the computational cost makes EGs an interesting way to approximate such densities. For instance, the MD simulations took 7 GPU-weeks for the Shh/heparin system and 12 GPU-weeks for the 14-3-3/QQJ-096 system, while the setup and computation for the EGs took 3–4 h of wall-clock time with a conventional CPU-based PC. However, there are also limitations that should be considered. There are two basic categories of deficiencies: first, those due to features of the real system that are missing in the model, and second, those due to inadequate configuration of the model.

Intraligand interactions fall into the first category of deficiencies. In the fragment-based screening used for the

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315–327

EGs, we neglect interactions between fragments. This can be problematic since we are looking at large, flexible ligands that carry charges and can have plenty of opportunities for such interactions, e.g., repulsion between same-sign charges, salt bridges, or π−cation interactions. It is clear that such interactions exist and that they can have an impact on the ligand structure.[69] A factor that could limit the severity of the effect of intramolecular interactions is that they have to compete with ligand−protein and ligand−water interactions and with the conformational entropy of the ligand.

Another example of a principal deficiency of the underlying model is the use of continuum electrostatics. This leads, e.g., to the neglect of structural water molecules, although such waters can be crucial for specific interactions at the surface of charged proteins.[70,71] A similar argument can be made for ions. Inclusion of fixed water molecules or ions in the protein structure is technically feasible in EG computations. For a first test, we used the Trp repressor interaction with a nucleotide ligand, where structural water molecules are known to mediate interactions.[72] In this case, the presence of the known structural water molecules had a negligible effect on the ligand distribution (Supporting Figure S5).

A further deficiency of our current EG implementation is the treatment of uncharged ligands. Since our focus was on charged ligands, this problem did not matter in the cases studied here. However, there are large, flexible biomolecules that do not carry formal charges and that are potentially interesting protein ligands. Since these uncharged ligands usually are still required to be soluble in aqueous solution, they will have to expose many hydrogen-bond donor and acceptor groups. For the same reason, protein surfaces typically are dominated by hydrogen-bond donors and acceptors and charged groups, too. Thus, we have to expect that hydrogen bonds play an important role in the interactions of proteins with such ligands. Since hydrogen bonds can be partly attributed to electrostatic interactions, these bonds can in principle be modeled by the proposed method. However, there is a crucial caveat resulting from the fact that hydrogen bonds are highly directional (in contrast to charge−charge interactions) and that the donor and acceptor groups are often flexible in that they can reorient easily, such as many hydroxyl groups in protein and carbohydrate side chains. This means that in order to pick up a clear electrostatic signal, the involved donors and acceptor groups on both the protein and the ligand should be in the right orientations, which is quite improbable because of their ability to reorient. For instance, in a complex of cyclomaltodextrin glucanotransferase with β-cyclodextrin (Supporting Figure S12H), the electrostatic interaction energy in the crystal complex amounts to $-26.3k_BT$ if we take the crystal structure of the complex and manually introduce the presumably correct hydrogen positions, but this affinity collapses to $-6.2k_BT$ if we do not take these precautions and choose the default hydrogen orientations of the GlyCAM force field. On the other hand, because the protein surface is cluttered with hydrogen-bond donors and acceptors, there will be ample competition for hydrogen-bond-mediated binding across the protein surface. Accordingly, in the case of the protein complex with β-cyclodextrin discussed above, large parts of the protein surface are sprinkled with elevated probability density (Supporting Figure S12G). Third, water itself is of course available for hydrogen-bonding interactions, which limits the free energy of protein−ligand binding in these cases. Taken together, these considerations indicate that the method studied in this work will in general not be sufficient for identifying binding regions on proteins for ligands devoid of formal charges but rich in hydrogen-bond donors and acceptors.

Furthermore, the current EG implementation does not explicitly account for hydrophobic interactions. It is unlikely that interactions of large, flexible ligands with proteins in aqueous solution are dominated by hydrophobic interactions because otherwise the binding partners would probably not be soluble. However, it is still possible that hydrophobic interactions between, say, a hydrophobic pocket of the protein and an alkyl group of the ligand, play a role in the interaction, so neglecting these interactions could lead to wrong probability densities. This problem could be addressed within the FFT-based EG computations by using complex numbers in the shape correlation (see Shape Complementarity), with one dimension modeling the shape and the other the hydrophobic interaction, similar to approaches that have been used in protein−protein docking.[73]

Yet another potential deficiency is the treatment of molecular flexibility. In the Kringle domain and $P2X_4$ systems the location of the probability density maximum was robust against conformational changes. However, it is easy to imagine cases where, for instance, charged groups are dislocated by a conformational change and fundamentally change the EG. If such conformational changes are known beforehand, they can be consistently included in the EG calculations on both the ligand side and the receptor side as demonstrated above. However, if not all of the relevant receptor conformations are known or if the protein−ligand interaction induces a new, unknown set of conformations, the distributions obtained from the EGs can be wrong. While this has not been an issue in the systems discussed above, it could be relevant for highly flexible proteins.

Finally, the FFT-based correlation computation used in Epitopsy treats the protein and ligand inconsistently: while the protein is modeled as a uniform medium with a low dielectric constant, the ligand is treated as point charges in continuum water. For flexible charged ligands, this approximation can be acceptable because they will be well-solvated and polarizable, but it could become inaccurate for ligands with more rigid structures or nonpolar regions.

The second category of deficiencies can be controlled by proper configuration of the model. For instance, if all of the relevant conformations of the ligand fragment and receptor are known, they should be included in the EG evaluation (eq 11), or at least the robustness of the EGs against conformational changes should be checked.

Another source of errors that can be controlled is the grid configuration. If the spatial or angular grids are too coarse, regions around the protein that contribute to the EG will be missed. In the present study, we used fragments with the maximum size of a disaccharide and a grid spacing of 0.8 Å, and the comparison with MD and experimental data showed that the results are reasonable for the given ligands and proteins. However, we expect that problems will arise with increasing ligand fragment size and ruggedness of the protein topography (Supporting Figure S6). For instance, the larger the ligand fragment and the deeper the protein pockets, the more difficult it will be to map the protein−ligand interaction on the angular and spatial grid because many fragment poses will lead to collisions and therefore be discarded. Another useful parameter in this context is the clash penalty δ (eq 1). A weaker penalty will increase the noise but also has the potential of making

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315−327

visible finer structures in the EG or probability density (Supporting Figure S4).

The correct charge of the ligand fragment is crucial, as shown in the epimerase example. However, the approach is robust against small variations in the charge distribution (see, e.g., Supporting Figure S7), allowing resource-intensive quantum-mechanics-based methods for charge assignment to be substituted with MM force field charges.

An important point that has not been addressed in this work is heterogeneous ligand composition. In the presented examples we could infer the location of larger ligands from fragment probability densities because the large ligand had a rather homogeneous composition, e.g., it was a heparin polysaccharide with negative charges on all of the disaccharides or a multibranched ligand with positively charged GCP and Lys groups. However, such large ligands may comprise subunits with different physicochemical characters. In this case, information on EGs for different ligand fragments have to be combined to infer likely locations of complete ligands. We are currently developing methods to postprocess sets of EGs for heterogeneous ligands in this sense.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.7b00413.

Supporting Table S1: Geometries of the glycosaminoglycans used as input for EGs. Supporting Table S2: Refined PDB structures. Supporting Figure S1: Initial placement of QQJ-096 in the MD simulations with 14-3-3/c-Raf. Supporting Figure S2: Initial placement of the heparin disaccharide in the MD simulations with sonic hedgehog. Supporting Table S3: MD−EG cluster distances. Supporting Figure S3: Effect of the number of rotations |Ω|. Supporting Figure S4: Effect of the grid resolution and penalty δ. Supporting Figure S5: Effect of explicit water molecules. Supporting Figure S6: Effect of the ligand size. Supporting Figure S7: Effect of the ligand charge distribution. Supporting Figure S8: Effect of shape complementarity and ligand charge distribution. Supporting Figure S9: Binding/unbinding events in the heparin/sonic hedgehog MD simulations. Supporting Figure S10: Distances between EG-predicted and crystallographically determined binding sites. Supporting Figure S11: 2D histograms of the ESP and EG vs MD probability density. Supporting Figure S12: Correlation with uncharged ligands (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: daniel.hoffmann@uni-due.de. Phone: +49 (0)201 183 4391. Fax: +49 (0)201 183 3437.

ORCID

Christian Ottmann: 0000-0001-7315-0315

Daniel Hoffmann: 0000-0003-2973-7869

Author Contributions

Conceived and designed the experiments: J.-N.G. and D.H. Performed the calculations: J.-N.G. Analyzed the data: J.-N.G., L.O., and D.H. Contributed computational tools: J.-N.G., C.W., J.N.D., and L.O. Contributed experimental data: A.G., C.S., and C.O. Wrote the paper: J.-N.G. and D.H.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft through Grant CRC 1093 to C.S. and A.G. (Subproject A1), J.-N.G., L.O., and D.H. (Subproject A7), and C.O. (Subproject B4).

■ REFERENCES

(1) Capila, I.; Linhardt, R. J. Heparin-Protein Interactions. *Angew. Chem., Int. Ed.* **2002**, *41*, 390−412.

(2) Coombe, D. R.; Kett, W. C. Heparan Sulfate-Protein Interactions: Therapeutic Potential Through Structure-Function Insights. *Cell. Mol. Life Sci.* **2005**, *62*, 410−424.

(3) Myszka, D. G.; Sweet, R. W.; Hensley, P.; Brigham-Burke, M.; Kwong, P. D.; Hendrickson, W. A.; Wyatt, R.; Sodroski, J.; Doyle, M. L. Energetics of the HIV gp120-CD4 Binding Reaction. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9026−9031.

(4) Willcox, B. E.; Gao, G. F.; Wyer, J. R.; Ladbury, J. E.; Bell, J. I.; Jakobsen, B. K.; van der Merwe, P. A. TCR Binding to Peptide-MHC Stabilizes a Flexible Recognition Interface. *Immunity* **1999**, *10*, 357−365.

(5) Gilles, P.; Wenck, K.; Stratmann, I.; Kirsch, M.; Smolin, D. A.; Schaller, T.; de Groot, H.; Kraft, A.; Schrader, T. High-Affinity Copolymers Inhibit Digestive Enzymes by Surface Recognition. *Biomacromolecules* **2017**, *18*, 1772−1784.

(6) Jiang, Q. Q.; Bartsch, L.; Sicking, W.; Wich, P. R.; Heider, D.; Hoffmann, D.; Schmuck, C. A New Approach to Inhibit Human β-Tryptase by Protein Surface Binding of Four-Armed Peptide Ligands With Two Different Sets of Arms. *Org. Biomol. Chem.* **2013**, *11*, 1631−1639.

(7) Yu, W.; Lakkaraju, S. K.; Raman, E. P.; MacKerell, A. D. Site-Identification by Ligand Competitive Saturation (SILCS) Assisted Pharmacophore Modeling. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 491−507.

(8) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(9) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Molecular Surface Recognition: Determination of Geometric Fit Between Proteins and Their Ligands by Correlation Techniques. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 2195−2199.

(10) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. Modelling Protein Docking Using Shape Complementarity, Electrostatics and Biochemical Information. *J. Mol. Biol.* **1997**, *272*, 106−120.

(11) Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: An FFT-Based Protein Docking Program With Pairwise Potentials. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 392−406.

(12) Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-Based Identification of Druggable 'Hot Spots' of Proteins Using Fourier Domain Correlation Techniques. *Bioinformatics* **2009**, *25*, 621−627.

(13) Pierce, B. G.; Wiehe, K.; Hwang, H.; Kim, B.-H.; Vreven, T.; Weng, Z. ZDOCK Server: Interactive Docking Prediction of Protein-Protein Complexes and Symmetric Multimers. *Bioinformatics* **2014**, *30*, 1771−1773.

(14) Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144−1149.

(15) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037−10041.

(16) Miranker, A.; Karplus, M. Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29−34.

(17) Lauri, G.; Bartlett, P. A. CAVEAT: A Program to Facilitate the Design of Organic Molecules. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 51−66.

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315−327

(18) Ngan, C. H.; Bohnuud, T.; Mottarella, S. E.; Beglov, D.; Villar, E. A.; Hall, D. R.; Kozakov, D.; Vajda, S. FTMAP: Extended Protein Mapping With User-Selected Probe Molecules. *Nucleic Acids Res.* **2012**, *40*, W271−W275.
(19) Rosenfeld, R.; Zheng, Q.; Vajda, S.; DeLisi, C. Flexible Docking of Peptides to Class I Major-Histocompatibility-Complex Receptors. *Genet. Anal.: Tech. Appl.* **1995**, *12*, 1−21.
(20) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(21) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.
(22) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622−655.
(23) Homeyer, N.; Horn, A. H. C.; Lanig, H.; Sticht, H. AMBER Force-Field Parameters for Phosphorylated Amino Acids in Different Protonation States: Phosphoserine, Phosphothreonine, Phosphotyrosine, and Phosphohistidine. *J. Mol. Model.* **2006**, *12*, 281−289.
(24) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.02; Gaussian, Inc.: Wallingford, CT, 2009; http://www.gaussian.com.
(26) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(27) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115*, 9620−9631.
(28) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(29) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; University of California: San Francisco, 2012; http://ambermd.org.
(30) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **2012**, *5*, 367.
(31) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(32) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(33) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(34) Páll, S.; Hess, B. A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures. *Comput. Phys. Commun.* **2013**, *184*, 2641−2650.
(35) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N\cdot\log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(36) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(37) Kaufman, L.; Rousseeuw, P. J. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods*; Dodge, Y., Ed.; North-Holland: Amsterdam, 1987; pp 405−416.
(38) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2017.
(39) Rousseeuw, P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53−65.
(40) Hyndman, R. J. Computing and Graphing Highest Density Regions. *Am. Stat.* **1996**, *50*, 120−126.
(41) *PyMOL*, version 1.8; Schrödinger, LLC: New York, 2016.
(42) Webb, B.; Sali, A. In *Protein Structure Prediction*; Kihara, D., Ed.; Springer: New York, 2014; Chapter 1, pp 1−15.
(43) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A. *Comparative Protein Structure Modeling Using MODELLER*; John Wiley & Sons: New York, 2001.
(44) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* **2007**, *35*, W522−W525.
(45) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665−W667.
(46) Singh, U. C.; Kollman, P. A. An Approach to Computing Electrostatic Charges for Molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.
(47) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.
(48) Gasteiger, J.; Marsili, M. A New Model for Calculating Atomic Charges in Molecules. *Tetrahedron Lett.* **1978**, *19*, 3181−3184.
(49) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity - A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(50) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*, 33.
(51) Wilms, C. Methods for the Prediction of Complex Biomolecular Structures. Ph.D. Dissertation, Universität Duisburg-Essen, Essen, Germany, 2013; http://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-34803/BIB_Thesis_Christoph_Wilms.pdf.
(52) Swinbank, R.; James Purser, R. Fibonacci Grids: ANovel Approach to Global Modelling. *Q. J. R. Meteorol. Soc.* **2006**, *132*, 1769−1793.
(53) González, Á. Measurement of Areas on a Sphere Using Fibonacci and Latitude-Longitude Lattices. *Math. Geosci.* **2010**, *42*, 49.
(54) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671−680.

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018, 58, 315−327

(55) Černý, V. Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm. *J. Optim. Theory Appl.* **1985**, *45*, 41−51.
(56) Whalen, D. M.; Malinauskas, T.; Gilbert, R. J. C.; Siebold, C. Structural Insights Into Proteoglycan-Shaped Hedgehog Signaling. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16420−16425.
(57) Jacobsson, I.; Lindahl, U.; Jensen, J. W.; Rodén, L.; Prihar, H.; Feingold, D. S. Biosynthesis of Heparin. Substrate Specificity of Heparosan *N*-Sulfate D-Glucuronosyl 5-Epimerase. *J. Biol. Chem.* **1984**, *259*, 1056−1063.
(58) Qin, Y.; Ke, J.; Gu, X.; Fang, J.; Wang, W.; Cong, Q.; Li, J.; Tan, J.; Brunzelle, J. S.; Zhang, C.; Jiang, Y.; Melcher, K.; Li, J.-p.; Xu, H. E.; Ding, K. Structural and Functional Study of D-Glucuronyl C5-Epimerase. *J. Biol. Chem.* **2015**, *290*, 4620−4630.
(59) Lindahl, U.; Kusche, M.; Lidholt, K.; Oscarsson, L. G. Biosynthesis of Heparin and Heparan Sulfate. *Ann. N. Y. Acad. Sci.* **1989**, *556*, 36−50.
(60) Jiang, Q.-Q.; Sicking, W.; Ehlers, M.; Schmuck, C. Discovery of Potent Inhibitors of Human β-Tryptase from Pre-Equilibrated Dynamic Combinatorial Libraries. *Chem. Sci.* **2015**, *6*, 1792−1800.
(61) Molzan, M.; Kasper, S.; Röglin, L.; Skwarczynska, M.; Sassa, T.; Inoue, T.; Breitenbuecher, F.; Ohkanda, J.; Kato, N.; Schuler, M.; Ottmann, C. Stabilization of Physical RAF/14-3-3 Interaction by Cotylenin A as Treatment Strategy for RAS Mutant Cancers. *ACS Chem. Biol.* **2013**, *8*, 1869−1875.
(62) Wu, T. P.; Padmanabhan, K.; Tulinsky, A.; Mulichak, A. M. The Refined Structure of the Epsilon-Aminocaproic Acid Complex of Human Plasminogen Kringle 4. *Biochemistry* **1991**, *30*, 10589−10594.
(63) Bodor, G.; Bednowitz, A. L.; Post, B. The Crystal Structure of ϵ-Aminocaproic Acid. *Acta Crystallogr.* **1967**, *23*, 482−490.
(64) Hattori, M.; Gouaux, E. Molecular Mechanism of ATP Binding and Ion Channel Activation in P2X Receptors. *Nature* **2012**, *485*, 207−212.
(65) Neale, C.; Hsu, J. C. Y.; Yip, C. M.; Pomès, R. Indolicidin Binding Induces Thinning of a Lipid Bilayer. *Biophys. J.* **2014**, *106*, L29−L31.
(66) Nemec, M.; Hoffmann, D. Quantitative Assessment of Molecular Dynamics Sampling for Flexible Systems. *J. Chem. Theory Comput.* **2017**, *13*, 400−414.
(67) Jones, S.; Shanahan, H. P.; Berman, H. M.; Thornton, J. M. Using Electrostatic Potentials to Predict DNA-Binding Sites on DNA-Binding Proteins. *Nucleic Acids Res.* **2003**, *31*, 7189−7198.
(68) Chen, Y. C.; Lim, C. Predicting RNA-Binding Sites from the Protein Structure Based on Electrostatics, Evolution and Geometry. *Nucleic Acids Res.* **2008**, *36*, e29.
(69) Lavery, R.; Maddocks, J. H.; Pasi, M.; Zakrzewska, K. Analyzing Ion Distributions Around DNA. *Nucleic Acids Res.* **2014**, *42*, 8138−8149.
(70) Materese, C. K.; Savelyev, A.; Papoian, G. A. Counterion Atmosphere and Hydration Patterns Near a Nucleosome Core Particle. *J. Am. Chem. Soc.* **2009**, *131*, 15005−15013.
(71) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 Å Resolution. *J. Mol. Biol.* **2002**, *319*, 1097−1113.
(72) Otwinowski, Z.; Schevitz, R. W.; Zhang, R. G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. Crystal Structure of trp Repressor/Operator Complex at Atomic Resolution. *Nature* **1988**, *335*, 321−329.
(73) Berchanski, A.; Shapira, B.; Eisenstein, M. Hydrophobic Complementarity in Protein-Protein Docking. *Proteins: Struct., Funct., Genet.* **2004**, *56*, 130−142.

DOI: 10.1021/acs.jcim.7b00413
J. Chem. Inf. Model. 2018. 58. 315−327

SUPPORTING INFORMATION

Locating Large, Flexible Ligands on Proteins

Jean-Noël Grad,[†] Alba Gigante,[‡] Christoph Wilms,[†] Jan Nikolaj Dybowski,[†] Ludwig Ohl,[†] Christian Ottmann,[¶] Carsten Schmuck,[‡] and Daniel Hoffmann[*,†]

[†]*Bioinformatics and Computational Biophysics, Faculty of Biology, University of Duisburg-Essen, Universitätstraße 7, 45141 Essen, Germany*

[‡]*Institute of Organic Chemistry, University of Duisburg-Essen, Universitätstraße 7, 45141 Essen, Germany*

[¶]*Laboratory of Chemical Biology, Department of Biomedical Engineering and Institute for Complex Molecular Systems, Eindhoven University of Technology, Den Dolech 2, 5612 AZ Eindhoven, The Netherlands*

E-mail: daniel.hoffmann@uni-due.de

Phone: +49 (0)201 183 4391. Fax: +49 (0)201 183 3437

Input structures

Supporting Table S1: Geometries of the glycosaminoglycans used as input for EGs, following the crystallographic definition of the glycosidic angles $\Phi(i)$ ($O_{5(i)}$–$C_{1(i)}$–$O_{X(i-1)}$–$C_{X(i-1)}$) and $\Psi(i+i)$ ($C_{1(i+1)}$–$O_{1(i+1)}$–$C_{X(i)}$–$C_{X-1(i)}$) for aldopyranoses.[1] The two heparin structures were determined from a 2 µs MD simulation trajectory partitioned in two clusters. The other angles were based on the GlcNS-IdoA linkage in PDB entry 5t03.[2] The glycosidic bonds in heparan sulfate are known to be highly flexible,[3] so we checked the EG of the heparosan against the EG of an alternative conformer with different angles still within the experimental error, and found no significant difference. All sugar rings have the conformation $^4C_1$.

Disaccharide unit	Linkage	Φ (°)	Ψ (°)
Heparin (cluster 1)	GlcNS(6S)-IdoA(2S)	63.7	95.0
Heparin (cluster 2)	GlcNS(6S)-IdoA(2S)	96.0	148.5
Desulfated heparin	GlcNS-IdoA	70.0	103.0
Heparan sulfate	GlcNS-GlcA	70.0	103.0
Heparosan	GlcNAc-GlcA	70.0	103.0
Heparosan (alt.)	GlcNAc-GlcA	100.0	79.0

heparin: 2-deoxy-4-*O*-methyl-2-(sulfoamino)-6-*O*-sulfonato-α-D-glucopyranosyl-(1→4)-1-*O*-methyl-2-*O*-sulfonato-α-L-idopyranuronate

desulfated heparin: 2-deoxy-4-*O*-methyl-2-(sulfoamino)-α-D-glucopyranosyl-(1→4)-1-*O*-methyl-α-L-idopyranuronate

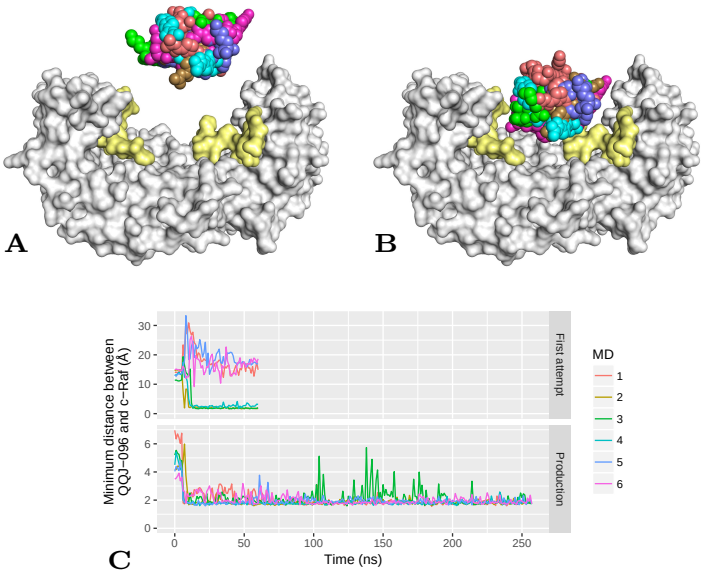
heparan sulfate: 2-deoxy-4-*O*-methyl-2-(sulfoamino)-α-D-glucopyranosyl-(1→4)-1-*O*-methyl-α-D-glucopyranuronate

heparosan: 2-(acetamido)-2-deoxy-4-*O*-methyl-α-D-glucopyranosyl-(1→4)-1-*O*-methyl-α-D-glucopyranuronate

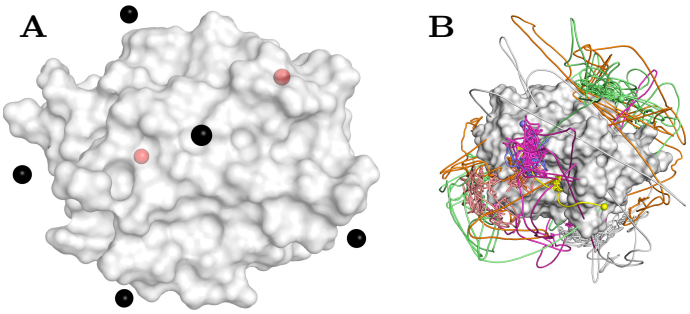
Supporting Table S2: PDB identifiers of crystal structures used as templates for refinement in Modeller and RMSD value between the templates and models.

Protein	Template	RMSD (Å)
Sonic Hedgehog	4c4n[4]	0.30
14-3-3ζ	4ihl[5]	0.44
C_5-epimerase	4pw2[6]	0.34
Trp repressor	1tro[7]	0.36

Supporting Figure S1: Initial placement of QQJ-096 in the MD simulations with 14-3-3/c-Raf. Based on the EGs obtained in Figure 5, the search space was reduced to the 14-3-3 pore where the two c-Raf peptides are located. Two conformations of QQJ-096 extracted from a 50 ns simulation in water were introduced in the 14-3-3/c-Raf simulation box and rotated to yield 6 different starting conditions. In a first series of 6 simulations, the minimal distance between QQJ-096 and the c-Raf peptides was above 10 Å (A), however in 3 runs the ligand did not reach the receptor within 50 ns (C, upper panel). In a second series of 6 simulations, the ligand was brought closer to the pore, with a minimal distance of 4–6 Å (B), resulting in immediate binding (C, lower panel). The plateau at the beginning of the simulations (C) corresponds to the NVT equilibration where the coordinates of heavy atoms are restrained.



Supporting Figure S2: Initial placement of the heparin disaccharide in the MD simulations with Sonic Hedgehog. (A) Sonic Hedgehog (semi-transparent white surface) and heparin di-saccharide centers (black spheres in front of the protein, red spheres behind the protein) scattered at random around the protein surface, with a minimal inter-atomic distance of 1.8 to 3.6 Å between the ligand and protein. (B) MD traces of the seven 500 ns simulations.

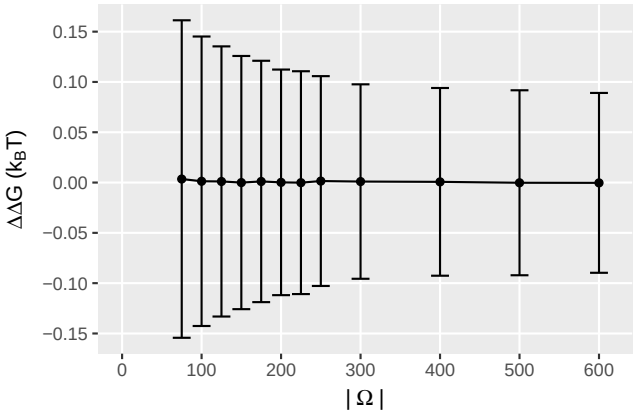


Supporting Table S3: Local minima in the EGs and MD-derived grids for scans with GCP and Lysine were detected and clustered as described in last section of Methods. For each MD medoid, the Euclidean distance to the nearest EG medoid is reported. For comparison: the length of GCP is about 10 Å, the length of Lys is about 8 Å.

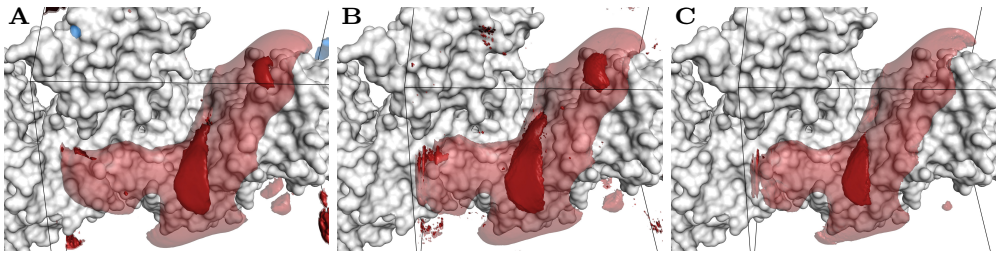
GCP		Lysine	
MD cluster size (%)	Distance (Å)	MD cluster size (%)	Distance (Å)
24	10	45	8.4
22	7.6	17	11
17	7.7	9.5	18
10	20	8.5	19
6.5	6.1	8.0	11
5.0	9.1	5.0	22
4.5	8.5	3.0	12
4.5	19	1.5	13
4.0	5.9	1.5	11
1.0	16	1.0	24
1.0	18	–	–

Effect of different choices of parameters

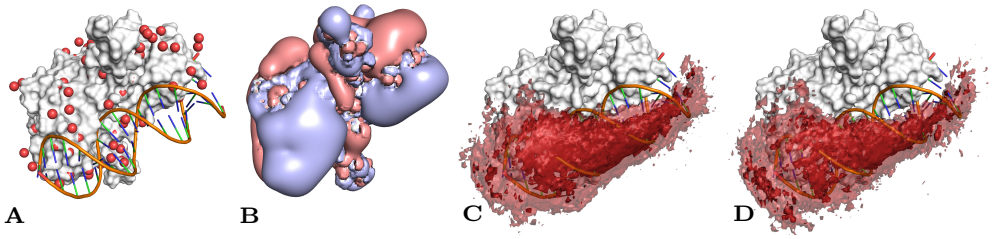
Supporting Figure S3: Effect of the number of rotations $|\mathbf{\Omega}|$ on the quality of EGs of C_5-epimerase with a CH_3O-GlcNAc-GlcA-GlcNS-OCH_3 trisaccharide as molecular probe at a resolution of 0.8 Å. We measure the energy difference between grid points of an EG calculated with $|\mathbf{\Omega}|$ rotations and a reference EG with $|\mathbf{\Omega}| = 800$, while limiting ourselves to grid points where the molecular probe made contact with the protein surface, i.e. where n angular states where allowed with $0 < n < |\mathbf{\Omega}|$. The average difference for any grid point in this region is close to zero with a 95% confidence interval decreasing as $|\mathbf{\Omega}|$ increases ($\pm 0.16 k_B T$ for $|\mathbf{\Omega}| = 75$, $\pm 0.10 k_B T$ for $|\mathbf{\Omega}| = 250$). The computation time increases linearly with $|\mathbf{\Omega}|$.



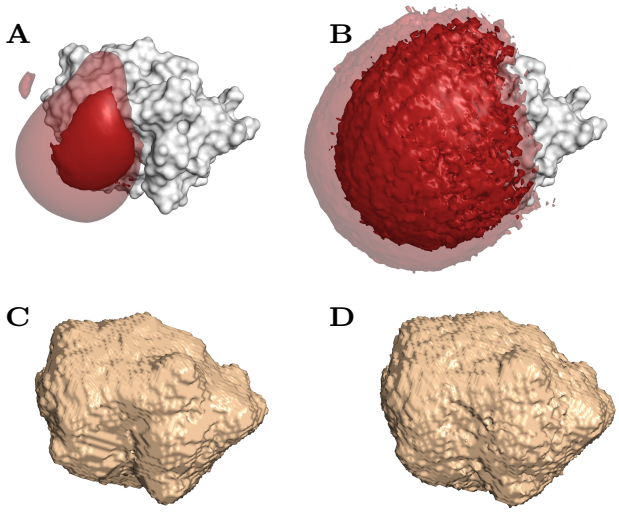
Supporting Figure S4: Effect of decreasing the grid size: EGs of C_5-epimerase scanned with the GlcNS-GlcA ligand at (A) 0.80 Å with penalty $\delta = -15$, (B) 0.40 Å with penalty $\delta = -3$, (C) 0.40 Å with penalty $\delta = -15$, all isosurfaces drawn at $\pm 1 k_B T$ (translucent blue/red) and $\pm 2 k_B T$ (solid blue/red) with black lines to represent the EG boundaries. The EG with $\delta = -3$ (A) shows slightly more detail but overall the same main structures as the EG with $\delta = -15$ (B). In contrast, details are lost if we improve resolution from 0.80 Å (A) to 0.40 Å (C) without adapting the penalty.



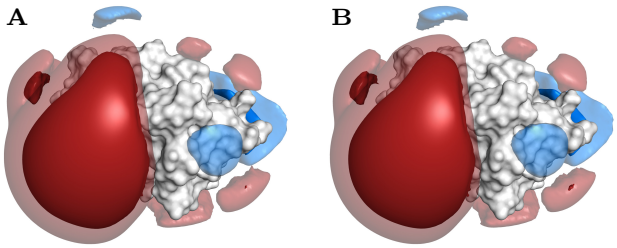
Supporting Figure S5: Effect of explicit water molecules: (A) Trp repressor (PDB entry 1tro [7]) with the crystallographic water molecules within 3 Å of the surface, (B) APBS grid without the water molecules, (C–D) epitopsy EGs using the crystallographic DNA fragment as molecular probe, with isosurfaces drawn at $-6k_{\mathrm{B}}T$ (translucent red) and $-8k_{\mathrm{B}}T$ (solid red), (C) without crystallographic waters and (D) with crystallographic waters.



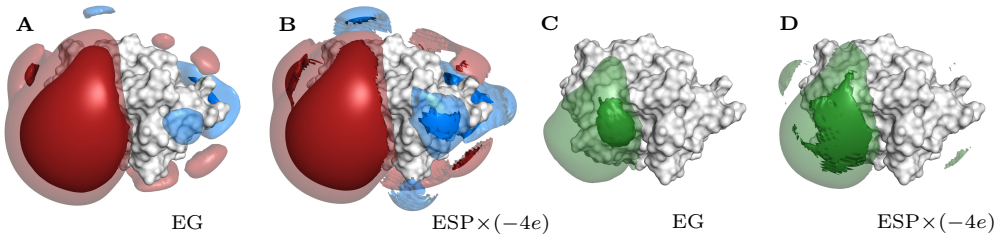
Supporting Figure S6: Effect of the ligand size: EGs of Sonic Hedgehog scanned with a heparin (A) dimer and (B) octamer, drawn at $-2k_{\mathrm{B}}T$ (translucent red) and $-4k_{\mathrm{B}}T$ (solid red), and ligand excluded volume (LEV) of the heparin (C) dimer and (D) octamer. The LEV in (D) is approximately two grid points larger than in (C), which corresponds to an 1.6 Å expansion. This offset can play an important role for EGs of protein surfaces with narrow clefts.



Supporting Figure S7: Effect of the ligand charge distribution: EGs of Sonic Hedgehog scanned with a heparin di-saccharide with partial charges obtained (A) from GlyCAM and (B) from delocalized formal charges ($-1/3$ on sulfate oxygens and $-1/2$ on carboxylate oxygens, zero on all other atoms), drawn at $\pm 1 k_{\mathrm{B}}T$ (translucent blue/red) and $\pm 2 k_{\mathrm{B}}T$ (solid blue/red).

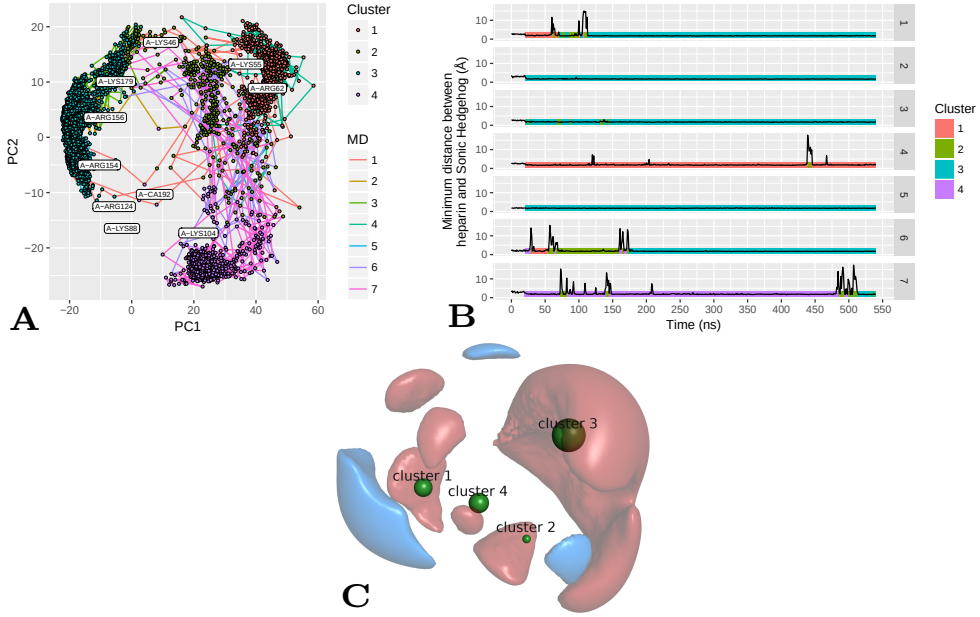


Supporting Figure S8: To study the contribution of shape complementarity and ligand charge distribution to EG and probability density, we compare the EG (A) and probability density (C) described in the main text with full correlation of shape complementarity and electrostatics based on charge distribution of heparin di-saccharide with a version of EG (B) and probability density (D) that neglects shape complementarity and focuses on the ESP alone. For the latter ESP-based EG the Shh protein surface was extended by 5 Å to approximate the ligand excluded volume of a heparin mono-saccharide, and correlated with a test point charge of $-4e$ (i.e. the total charge of the heparin di-saccharide). (A) EG isosurfaces at $\pm 1 k_{\mathrm{B}}T$ (translucent blue/red) and $\pm 2 k_{\mathrm{B}}T$ (solid blue/red), merging across two population-weighted conformations of a heparin di-saccharide. (B) ESP$\times(-4e)$ isosurfaces at $\pm 1 k_{\mathrm{B}}T$ (translucent blue/red) and $\pm 2 k_{\mathrm{B}}T$ (solid blue/red) using the test charge. (C) EG based probability density of heparin di-saccharide drawn around 20% HDR (solid green) and 30% HDR (translucent green). (D) ESP$\times(-4e)$ of a test charge drawn around 5% HDR (solid green) and 10% HDR (translucent green).

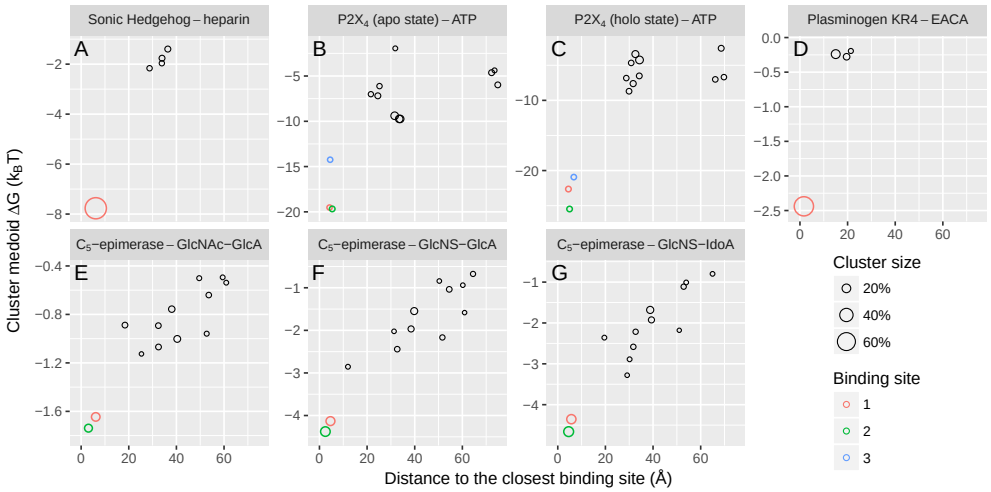


MD trajectory and EG clustering

Supporting Figure S9: Frequent binding sites on the surface of Sonic Hedgehog were extracted from the MD simulations using inter-residues, distance-based PCA.[8] The distances were computed as the minimum distance between basic amino acids in sonic hedgehog and negatively-charged groups ($R-SO_3^-$, $R-CO_2^-$) in heparin MD. Silhouette-validated PAM clustering of the binding sites was conducted on the first 6 principal components of the PCA matrix and yielded 4 clusters. (A) The 7 heparin trajectories projected into PCA space as dots colored by cluster. (B) Binding/unbinding events (spikes at 10–15 Å) are necessary for heparin to sample other binding sites. The still frames with no cluster information correspond to the 20 ns NVT equilibration with position restraints. (C) The 4 clusters are depicted as green spheres with log-radii proportional to the cluster population, superimposed to the EG isosurfaces (Figure 2B, $\pm 1 k_B T$).

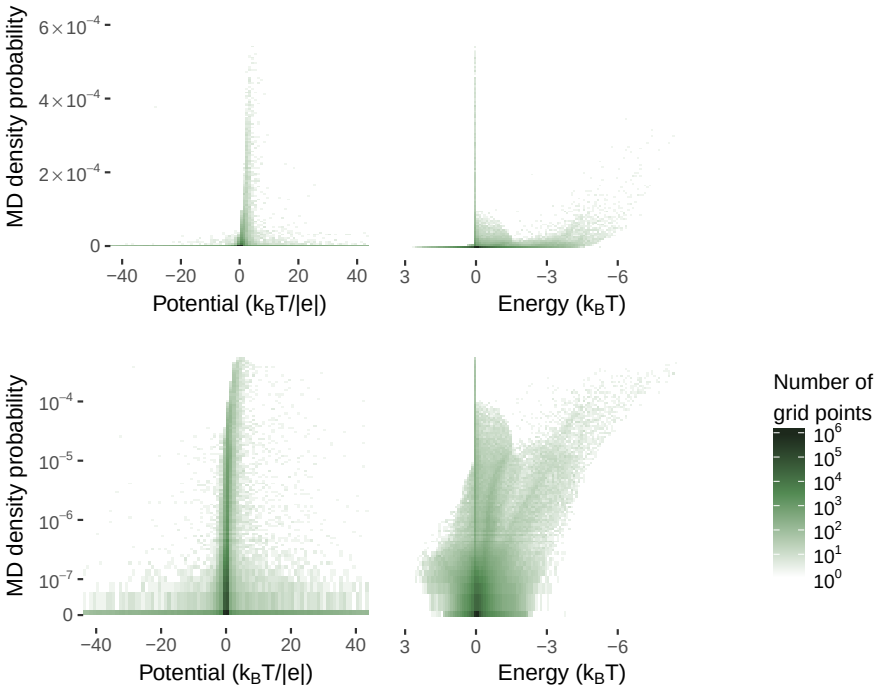


Supporting Figure S10: For each of the protein-ligand complexes the distance between cluster medoids (see Methods, section Quantitative comparison of EGs with experimental ligand positions) and center of crystallographic ligand position (x-axes) is plotted against the EG energy value of the corresponding medoid (y-axes). The plot shows that the clusters with the lowest energy are always the closest to the binding site(s). The x-coordinates of the colored circles correspond to d_{X-EG}. The circle size is proportional to the cluster size, i.e. the number of simulated annealing chains that converged to the same cluster. For the all proteins except $P2X_4$ the cluster closest to the binding site is also the largest cluster.



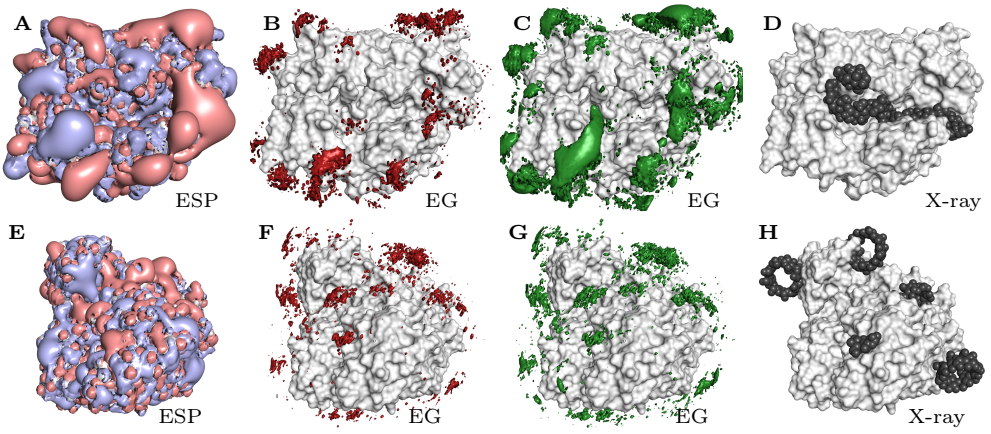
ESP/EG correlation histograms

Supporting Figure S11: Correlation between the MD probability density and the ESP grid and EG of Sonic Hedgehog with a heparin di-saccharide as molecular probe at a resolution of 0.8 Å. Attractive regions have positive potential in ESP and negative EG value. Binning the average MD occupancy versus the electrostatic potential, respectively, the energy in 2D histograms shows a higher correlation between EG and MD. In both histograms, the vertical line centered at zero corresponds to grid points far away from the protein where the heparin was not making contact with the protein and was randomly navigating in water. The horizontal baseline in ESP corresponds to zones of high potential values never sampled in MD due to the van der Waals exclusion zone. The ESP distribution is slightly skewed towards grid points of positive energies, but the correlation is barely measurable. The Epitopsy distribution has two visible plumes in regions of negative energies. The second plot shows the MD occupancy on the log scale with zero occupancy indicated by 0.



Uncharged ligands

Supporting Figure S12: Robust correlation requires electrical charges on the molecular probe. Although the charge grid is fine enough to capture the dipole moment of hydrogen bond donors and acceptors, this class of interaction is very sensitive to orientation and distance. Experiments with 4-α-glucanotransferase/amylose (A-D) and cyclomaltodextrin glucanotransferase/cyclodextrin (E-H) complexes show that binding regions cannot be accurately predicted in Epitopsy. (A) ESP isosurfaces of 4-α-glucanotransferase at $\pm 1 k_{\mathrm{B}}T/|e|$ (blue/red). (B) EG isosurfaces at $\pm 1 k_{\mathrm{B}}T$ (blue/red) using a glucose di-saccharide. (C) EG based probability density of the glucose di-saccharide drawn around 1% HDR (green). (D) crystal structure (PDB entry 5jiw[9]). (E) ESP isosurfaces of cyclomaltodextrin glucanotransferase at $\pm 0.1 k_{\mathrm{B}}T/|e|$ (blue/red). (F) EG isosurfaces at $\pm 0.1 k_{\mathrm{B}}T$ (blue/red) using a cyclodextrin. (G) EG based probability density of the cyclodextrin drawn around 0.1% HDR (green). (H) crystal structure (PDB entry 1d3c[10]).



References

(1) IUPAC, Symbols for Specifying the Conformation of Polysaccharide Chains. *Eur. J. Biochem.* **1983**, *131*, 5–7.

(2) Xu, Y.; Moon, A. F.; Xu, S.; Krahn, J. M.; Liu, J.; Pedersen, L. C. Structure Based Substrate Specificity Analysis of Heparan Sulfate 6-O-Sulfotransferases. *ACS Chem. Biol.* **2017**, *12*, 73–82.

(3) Khan, S.; Fung, K. W.; Rodriguez, E.; Patel, R.; Gor, J.; Mulloy, B.; Perkins, S. J. The Solution Structure of Heparan Sulfate Differs From That of Heparin: Implications for Function. *J. Biol. Chem.* **2013**, *288*, 27737–27751.

(4) Whalen, D. M.; Malinauskas, T.; Gilbert, R. J. C.; Siebold, C. Structural Insights Into Proteoglycan-Shaped Hedgehog Signaling. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16420–16425.

(5) Molzan, M.; Kasper, S.; Röglin, L.; Skwarczynska, M.; Sassa, T.; Inoue, T.; Breitenbuecher, F.; Ohkanda, J.; Kato, N.; Schuler, M.; Ottmann, C. Stabilization of Physical RAF/14-3-3 Interaction by Cotylenin A as Treatment Strategy for RAS Mutant Cancers. *ACS Chem. Biol.* **2013**, *8*, 1869–1875.

(6) Qin, Y.; Ke, J.; Gu, X.; Fang, J.; Wang, W.; Cong, Q.; Li, J.; Tan, J.; Brunzelle, J. S.; Zhang, C.; Jiang, Y.; Melcher, K.; Li, J.-p.; Xu, H. E.; Ding, K. Structural and Functional Study of D-Glucuronyl C5-Epimerase. *J. Biol. Chem.* **2015**, *290*, 4620–4630.

(7) Otwinowski, Z.; Schevitz, R. W.; Zhang, R. G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. Crystal Structure of trp Repressor/Operator Complex at Atomic Resolution. *Nature* **1988**, *335*, 321–329.

(8) Ernst, M.; Sittel, F.; Stock, G. Contact- and Distance-Based Principal Component Analysis of Protein Dynamics. *J. Chem. Phys.* **2015**, *143*, 244114.

(9) Roth, C.; Weizenmann, N.; Bexten, N.; Saenger, W.; Zimmermann, W.; Maier, T.; Sträter, N. Amylose Recognition and Ring-Size Determination of Amylomaltase. *Sci. Adv.* **2017**, *3*, e1601386.

(10) Uitdehaag, J. C.; Kalk, K. H.; van Der Veen, B. A.; Dijkhuizen, L.; Dijkstra, B. W. The Cyclization Mechanism of Cyclodextrin Glycosyltransferase (CGTase) as Revealed by a Gamma-Cyclodextrin-CGTase Complex at 1.8 Å Resolution. *J. Biol. Chem.* **1999**, *274*, 34868–34876.

2.2 Discussion

2.2.1 Molecular modeling

2.2.1.1 Sonic Hedgehog

Most crystal structures available for Shh feature a ligand. Two structures feature Shh with glycosaminoglycans: PDB entries 4c4n and 4c4m (Whalen et al. 2013), where heparin resp. chondroitin-4-sulfate are bound to the two-calcium state of murine Shh with a shortened *N*-terminus. The Ca2 state is typical for Shh in blood plasma, where calcium is found in a higher concentration than in the cytoplasm (Clapham 2007). Two structures feature Shh in the apo state: PDB entry 3m1n (Pepinsky et al. 2000) with Shh in the Ca0 state, and PDB entry 6pjv (Bonn-Breach et al. 2019), with Shh in the Mg1 state with an atypical coordination geometry for the zinc cation.

For consistency of the analysis between the X-ray structure and *in silico* predictions, Shh was modeled after the structure in PDB entry 4c4n. The flexible terminal loops, including the CW domain, were not restored to their full wild-type length to limit flexibility during MD.

2.2.1.2 Generation of a conformer ensemble for heparin

Heparin is a polymer of D-glucosamine *N*-sulfate (GlcNS) and L-iduronate (IdoA), with a random sulfation pattern on position 2 of IdoA and on position 6 of GlcNS. Heparin was modeled as a methyl-blocked di-saccharide[2] CH_3O-GlcNS(6S)-IdoA(2S)-OCH_3 in TLEaP with the default $^4C_1$ chair conformation (Ernst et al. 1998) for both saccharides. Initial dihedral angles at the glycosidic bond were obtained from PDB entry 4c4n (Whalen et al. 2013).

The di-saccharide conformational dynamics were sampled by MD in water for 2 µs. Although the iduronate ring can exist in both the $^4C_1$ and $^2S_0$ conformations under this sulfation pattern (Muñoz-García et al. 2013; Hsieh et al. 2016), ring puckering was not observed during the simulation. This is consistent with previously reported simulations of IdoA(2S) mono-saccharide that suggested $^2S_0$ was an intermediate in the $^4C_1 \leftrightarrows {}^1C_4$ interconversion with an exchange rate of the order of 10 µs (Sattelle et al. 2010).

Clustering of the MD trajectory revealed two clusters with good mixing (Figure 2.1). The main difference resided in the dihedral angle at the glycosidic bond; the Φ and Ψ angles[3] were 63.7°,95.0° for cluster 1 and 96.0°,148.5° for cluster 2 (Table S1 in Work 1). Experimental data on dihedral angles for free and protein–bound heparin fragments shows large fluctuations around mean values (Mulloy and Forster 2000, Figure 6B; Khan et al. 2013, Table 2), which could be explained by the presence of multiple low-energy equilibrium values. Both conformers were used as input for Epitospy, weighted by their respective cluster frequency, using partial charges from the GROMACS topology.

[2] IUPAC nomenclature: 2-deoxy-4-*O*-methyl-2-(sulfoamino)-6-*O*-sulfonato-α-D-glucopyranosyl-(1→4)-1-*O*-methyl-2-*O*-sulfonato-α-L-idopyranuronate.

[3] Φ and Ψ angles follow the crystallographic definition of the glycosidic angles $\Phi(i)$ ($O_{5(i)}$–$C_{1(i)}$–$O_{X(i-1)}$–$C_{X(i-1)}$) and $\Psi(i+i)$ ($C_{1(i+1)}$–$O_{X(i+1)}$–$C_{X(i)}$–$C_{X-1(i)}$) for aldopyranoses (IUPAC 1983).

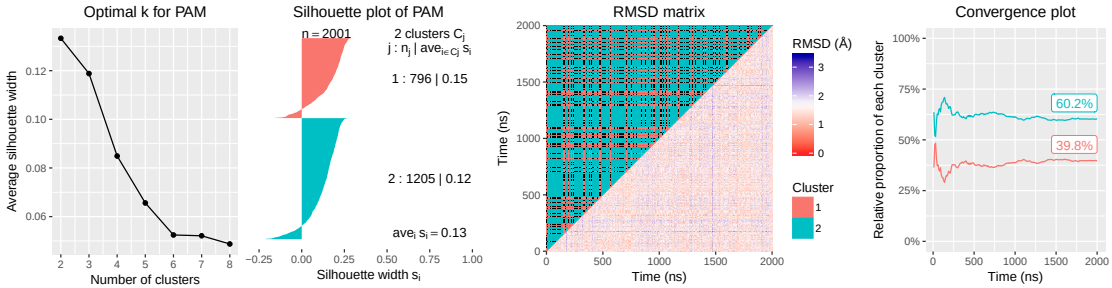


Figure 2.1: Clustering of the heparin di-saccharide MD simulation using the PAM algorithm. The silhouette plot shows no improvement in selecting more than two clusters. The two clusters frequently inter-convert according to the RMSD plot. The relative proportion of clusters stabilizes at 60% and 40% according to the convergence plot.

2.2.2 Modeling the interaction

2.2.2.1 MD simulations

For a fair comparison between MD and EG, it is crucial to sample the interaction energy landscape uniformly. Unfortunately, running an MD simulation with a long simulation time seldom results in a uniform sampling, as the system can be trapped in a local minimum with a high energy barrier. This situation is commonly referred to as the MD sampling problem (Zuckerman 2011; Maximova et al. 2016). Multiple enhanced sampling techniques have been designed to overcome energy barriers (Bernardi et al. 2015), although multiple-copy dynamics (Perez et al. 2016; Yan et al. 2020) offers a simpler and more straightforward setup: running multiple short trajectories with different initial conditions and concatenating them into a long trajectory. Using this strategy, the Shh:heparin system was simulated for a total of 3.5 µs.

In order to quantify the frequency of binding/unbinding events and rank the most favorable binding sites, the trajectories were clustered based on the contact distance between the protein and ligand. While coordinate-based clustering methods are a standard procedure for rigid proteins interacting with small ligands, distance-based clustering methods are more advantageous when flexibility is present (Ernst et al. 2015). The metric measures the distance from a basic residue (Arg, Lys) in Shh to the nearest anion ($R-SO_3^-$, $R-CO_2^-$) in the heparin di-saccharide[4]. To further reduce the dimensionality of the problem, principal component analysis (PCA) was used to extract the 6 dimensions that contributed most to the variability of the dataset. Silhouette-validated PAM clustering of these 6 principal components produced four clusters (Figure S9**A** in Work 1). Converting these clusters back to Cartesian coordinates revealed that cluster number 3 was located inside the blob of highest energy in the EG (Figure S9**C** in Work 1), where residue Arg^{156} is located. While the ligand could freely move between the other clusters, once it reached cluster 3, it was unable to detach itself from the binding site within the 0.5 µs of the individual simulations (Figure S9**B** in Work 1, blue highlight).

[4] An intuitive understanding of this metric can be gained by relating the inter-residue contact distance r of the cation–anion pair to the electrostatic energy U_E of two point charges in a medium with constant permittivity, where $U_E \propto r^{-1}$. Contact distances contribute to a cluster proportionally to the Coulomb energy.

2.2.2.2 The role of shape correlation

Plotting the MD probability density against the EG and electrostatic potential (ESP) (Figure 3 in Work 1) clearly demonstrates the importance of shape correlation in the formulation of the correlation energy. A qualitative picture is obtained by drawing the corresponding isosurfaces (Figure S8**C-D** in Work 1), which show a larger spatial variance for the point-like probe.

To better visualize the discrepancy between ESP and EG, the grid points in Figure 3 in Work 1 were binned and plotted in a 2D histogram (Figure S11 in Work 1, top row). The histograms reveal a correlation between EG values and MD occupancies (Figure S11 in Work 1, top right), which appears to be linear in the log plot (Figure S11 in Work 1, bottom right). This observation is consistent with the Boltzmann distribution. Conversely, the ESP distribution isn't remarkably correlated to the MD occupancies (Figure S11 in Work 1, bottom left).

To understand the origin of the difference between ESP and EG, it is helpful to plot one against the other (Figure 2.2, left plot). A pattern starts to emerge, which is more visible when the histogram is split in three histograms based on the number of allowed rotational states $|\omega|$, i.e. the number of ligand rotations for which the quantity $\mathbf{f}^{\mathrm{vdw}}_{C_{\alpha,\beta,\gamma}}$ in Equation 3 in Work 1 is non-negative (Figure 2.2, three plots on the right). The region $|\omega| = 0$ contains 3.2% of the grid points and corresponds to the ligand excluded volume. The EG is zero while the ESP has a high dynamic range, with several grid points reaching a potential of ±40 $k_B T/|e|$ (visible in Figure S11 in Work 1, bottom left), because positions in that region are accessible to water molecules but inaccessible to the ligand due to van der Waals repulsion.

The region $|\omega| = |\Omega|$ contains 91.4% of the grid points and corresponds to positions far enough from the protein surface to allow full rotation of the ligand. This region is characterized by a linear correlation between the ESP and EG. Fitting a linear model yields the equation $\hat{y} = -0.06\, k_B T/|e| + 0.94\, x/q^{\mathrm{ligand}}$ with $q^{\mathrm{ligand}} = -4\,|e|$ for the heparin di-saccharide probe. This model is close to the expected result for a point charge, i.e. $y = x/q^{\mathrm{ligand}}$. The region $0 < |\omega| < |\Omega|$ contains 5.4% of the grid points and corresponds to positions where the ligand makes physical contact with the protein surface. This region features a strong deviation from linearity due to the treatment of shape complementarity.

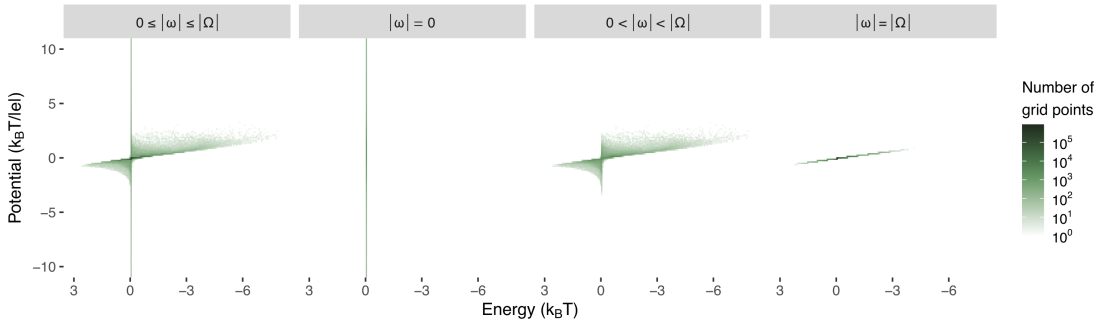


Figure 2.2: 2D histogram of the ESP and EG values for Shh with a heparin di-saccharide as molecular probe at a resolution of 0.8 Å and $|\Omega| = 1200$ (plot $0 \le |\omega| \le |\Omega|$), with decomposition based on the number of available rotational states ω (plots $|\omega| = 0$, $0 < |\omega| < |\Omega|$, $|\omega| = |\Omega|$). The set ω is a subset of Ω.

This experiment shows that protein electrostatics, on its own, is not an adequate tool to investigate protein–ligand energetic landscapes. The striking difference between the ESP and EG correlations in Figure 3 in Work 1 is mostly due to the region $|\omega| < |\Omega|$, which accounts for less than 9% of the grid points. In this region the ligand can no longer be accurately described by a point charge. This limitation of the PB model is addressed by the rigid body correlation method, which uses a simple charge and van der Waals representation of the ligand as the probe. The resulting EG has an intuitive interpretation and can be directly compared to computationally-expensive MD simulations. Rigid body correlation also works with zwitterionic probes, which cannot be represented in a point-like model.

2.2.2.3 Domain of validity for rigid correlation

Interaction model approximation

For the correlation to be meaningful, both the receptor and probe need to carry at least one electrical charge. Zwitterionic probes, although electrically neutral, also qualify (Figure 6 in Work 1). Probes with no charges produce a weak correlation signal indistinguishable from noise (Figure S12 in Work 1). Conversely, proteins with high charge density, such as nucleosome, are poor candidates due to the limitations of the mean-field-based PB equation used in APBS (Kirmizialtin et al. 2012). Other types of interactions are not modeled by Epitopsy. Section 1.1.3 "Diversification of the energetic model" gives an overview of other implementations of the rigid body correlation method that are not limited to the electrostatic potential.

Continuum model approximation

The mean-field based PB equation models the solvent as a continuum and introduces implicit counter-ions as non-interacting point charges. This assumption breaks down for highly charged biomolecules, such as nucleic acids (Kirmizialtin et al. 2012) and their receptors, leading to unphysical ions concentration close to the surface. The size-modified Poisson–Boltzmann (SMPB) model is an extension of the PB model where the counter-ions are interacting particles with a non-zero volume. Multiple versions of the SMPB equation were formulated (Borukhov et al. 1997; Chu et al. 2007; Kirmizialtin et al. 2012), resulting in more realistic ESP grids for DNA duplexes with salt concentrations up to 0.15 mol l^{-1} (Chu et al. 2007). The SMPB method was added to APBS in 2007 (Chu et al. 2007) but later removed due to numerical instability[5].

Similarly, the formation of an ionic atmosphere (Chu et al. 2008; Kirmizialtin et al. 2012; Lavery et al. 2014; Gebala et al. 2016) around these highly charged proteins poses the question of the inclusion of tightly and loosely bound salts and water molecules, which are usually visible in X-ray diffraction structures but not always fully or correctly captured (Carugo and Bordo 1999; Handing et al. 2018; Zheng et al. 2017). Water molecules in particular are known to extend hydrogen bonds (Williams et al. 1994; Janin 1999; Jiang et al. 2005) and salt-bridges (Papoian et al. 2003; Sabarinathan et al. 2011) between proteins and their ligands, but can also bridge two ions in the protein, in which case they can be difficult to displace by the ligand due to both enthalpic and entropic factors (Kubinyi 2007; Klebe 2015).

[5] GitHub repository Electrostatics/apbs-pdb2pqr: method SMPB was disabled on March 7, 2017 (commits c1c658d, 286292c). Last visited on November 28, 2019.

Crystallographic water molecules are crucial to understand protein binding to ligands, but are also challenging to identify and properly model (Bodnarchuk 2016; Jukič et al. 2017). To evaluate the impact of such water molecules, two EGs were calculated for the RNA-binding Trp-repressor protein, using the protein structure with and without crystallographic water molecules (Figure S5 in Work 1). Apart from extending the protein molecular surface by the diameter of a water molecule, the presence or absence of water did not play a significant role.

Inhomogeneous dielectric medium

Several families of proteins interacting with charged ligands are found on the cell membrane. Although Epitopsy in its current form can only treat proteins in a homogeneous solvent, the EGs computed for the ATP-gated $P2X_4$ channel show Epitopsy could still be used with reasonable accuracy for transmembrane proteins, even though the lipid bilayer was modeled as saline water. Such an approximation can be justified by the following two factors: the ATP binding site is located far away from the lipid bilayer, and the protein surface in contact with the lipid bilayer is mostly apolar, therefore contributing minimally to the EG. This is however not true for all transmembrane proteins, and applications of Epitopsy to such systems should be decided on a case-by-case basis. For systems where the binding site is close to the lipid bilayer, the electric charges from the lipid tails and the potential difference between the cell interior and exterior cannot be neglected anymore. APBSmem (Callenberg et al. 2010) could be used instead of APBS as the electrostatics solver backend to properly model the change of dielectric, although this option was not explored in Epitopsy due to the difficulty to compare predictions to accurate protein MD simulations at the lipid/water interface.

Protein and ligand flexibility

The rigid nature of complementarity scanning can be seen as a limitation of the method, since many biologically relevant systems exhibit some level of flexibility. This shortcoming is traditionally solved from the ligand side by either selecting a consensus ligand pose that optimizes the prediction against a collection of protein targets (Mottarella et al. 2014), or by using a coarser grid for the ligand (Vakser 1995; Vakser 1996). Alternatively, one can employ a two-stage approach where the predicted binding sites are extracted from the EG and used as input for a finer exhaustive search (Gabb et al. 1997) or for a heuristic search with a different algorithm that can treat flexible groups (Li et al. 2003; Sternberg et al. 1998), or both for a three-stage approach (Meyer et al. 1996). From the receptor side, one approach consists in generating perturbed structures using a rotamer library and selecting the structure with highest affinity for the ligand (Grove et al. 2013).

Using an ensemble of ligand poses for ε-aminocaproic acid and protein conformations for plasminogen Kringle Domain 4, Epitopsy was able to recover the position of the binding site with high accuracy via a simple arithmetic mean (Figure 6**G** in Work 1). This suggests that the complementarity scanning method can account for flexibility simply by merging EGs together, without the need to select or filter conformers. In the case where the conformational ensemble is generated by a method that also produces relative frequencies, this information can be incorporated in the averaging step using a weighted arithmetic mean. For example, the frequency of the two heparin di-saccharide conformations was used as weights to calculate

the EGs of the Shh:heparin system.

Conformational averaging of EGs is not a viable strategy for proteins with extreme plasticity, such as the ATP-gated $P2X_4$ channel, where the binding site moves by several grid points between the apo and holo states. In this situation, Epitopsy was able to correctly locate the binding site in the bound and unbound states separately (Figure 7**C-F** in Work 1 and Figure S10**B-C** in Work 1). This system also reached the limit of what Epitopsy could achieve for buried binding sites: the features of the EG in the bound state are dominated by three grid points, one per binding site (Figure 7**F** in Work 1).

3 Chapter
Epitopsy-directed MD simulations

14-3-3 proteins are ubiquitous signal transducing adaptor proteins that regulate signaling pathways by altering the activity or function of signaling proteins through reversible binding (Mackintosh 2004). The family of 14-3-3 proteins contains 153 known isoforms across unicellular and multicellular organisms (Rosenquist et al. 2000), with seven in humans (Aitken 2002). Key client proteins are found in the rapidly accelerated fibrosarcoma (RAF) kinase family: RAF proto-oncogene serine/threonine-protein kinase (c-Raf), serine/threonine-protein kinase A-Raf (A-Raf) and serine/threonine-protein kinase B-Raf (B-Raf). These three enzymes self-assemble into homo-dimers and hetero-dimers (Rushworth et al. 2006; Freeman et al. 2013; Varga et al. 2017; Rajakulendran et al. 2009). 14-3-3 plays a central role in the regulation of c-Raf (Khazak et al. 2007), c-Raf:Ras complexes (McCubrey et al. 2006) and c-Raf:B-Raf complexes (Rushworth et al. 2006; Varga et al. 2017).

B-Raf and c-Raf are part of the Ras-Raf-MEK-ERK signaling pathway (McCubrey et al. 2007). Disruption of this pathway is linked to numerous health disorders (Kim and Choi 2010; Zebisch and Troppmair 2006; Montagut and Settleman 2009). This pathway is relevant to oncology, where B-Raf (Khazak et al. 2007; Shepherd et al. 2010; Huang et al. 2013; Rajakulendran and Adam 2014; Hertzman Johansson and Egyhazi Brage 2014; Ascierto et al. 2012) and c-Raf (McPhillips et al. 2006; Heidorn et al. 2010; Blasco et al. 2011) are well-established therapeutic targets.

14-3-3 is also relevant to oncology (Hermeking 2003; Wilker and Yaffe 2004; Aghazadeh and Papadopoulos 2016) and is considered a potential therapeutic target (Yang et al. 2012; Matta et al. 2012; Neal and Yu 2010; Cui et al. 2020). Inhibitors were developed to destabilize the 14-3-3 dimer interface through phosphorylation (Woodcock et al. 2015) and steric effects (Ehlers et al. 2018), however the loss of 14-3-3 function affects all client proteins rather indiscriminately.

Recently, an alternative strategy targeting the protein–protein interaction (PPI) between 14-3-3 and specific client proteins emerged (Kaplan et al. 2017; Kaplan and Fournier 2017). This paradigm shift opens up new perspective for rational ligand design, as the PPI can be modulated through the inhibition or stabilization of specific 14-3-3:c-Raf complexes (Mori et al. 2013; Bartel et al. 2014; Hartman and Hirsch 2017), instead of 14-3-3 or c-Raf individually. Although inhibition is the most frequent choice in PPI modulation, for example using tweezers (Bier et al. 2013), stabilization is another possibility, for example with natural molecules cotylenin A and fusicoccin A (Molzan et al. 2013; Milroy et al. 2015).

This chapter investigates the 14-3-3:c-Raf PPI stabilization by polymer QQJ-096 via all-atom MD simulations. It is a direct continuation of the preliminary results and observations made in

Work 1, section "14-3-3 Protein and Polycationic Supramolecular Ligand." This project provides a general procedure for planning MD simulations of large proteins with highly flexible ligands using EGs. Contrary to the previous chapter, where quantitative agreement between MD and EG could be shown due to the simplicity of the Shh:heparin system, the flexible nature of the 14-3-3:c-Raf:QQJ-096 system made MD convergence an unrealistic goal to achieve. Instead, two rigid and charged fragments of the stabilizer were used to narrow down the putative binding site to a small area of the protein surface. Simulations were carried out with QQJ-096 initially placed at that location with a random orientation, and a prediction of the most probable glutamate and aspartate binding partners was obtained.

This chapter includes the following work: Alba Gigante, Jean-Noël Grad, Jeroen Briels, Maria Bartel, Daniel Hoffmann, Christian Ottmann, Carsten Schmuck, "A new class of supramolecular ligands stabilizes 14-3-3 protein–protein interactions by up to two orders of magnitude". *Chemical Communications* 2019, 55(1), 111–114. doi:10.1039/C8CC07946C.
Reproduced from reference Gigante et al. 2019 with permission from The Royal Society of Chemistry.
https://pubs.rsc.org/en/content/articlelanding/2019/CC/C8CC07946C

ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2019, 55, 111

Received 4th October 2018,
Accepted 26th November 2018

DOI: 10.1039/c8cc07946c

rsc.li/chemcomm

A new class of supramolecular ligands stabilizes 14-3-3 protein–protein interactions by up to two orders of magnitude†

A. Gigante, [a] J.-N. Grad, [b] J. Briels,[ac] M. Bartel,[ac] D. Hoffmann, *[b] C. Ottmann *[ac] and C. Schmuck *[a]

We report the first supramolecular stabilizers of the interaction between 14-3-3ζ and two of its effectors, Tau and C-Raf, which are involved in neurodegenerative diseases and proliferative signal transduction, respectively. These supramolecular ligands open up an opportunity to modulate functions of 14-3-3 with these effectors.

Due to the key role of protein–protein interactions (PPIs) in the homeostasis of biological systems, their dysfunction is involved in many pathologies and diseases like cancer, neurodegeneration or viral/parasitic infections.[1,2] In the last years a number of successful examples of molecules that are able to modulate PPIs have been carried on into the clinic.[3,4] Thus, protein surface recognition has become a promising tool to modulate protein function by disrupting or stabilizing PPIs.[5]

The human 14-3-3 protein family is one of the regulatory elements in intracellular signalling pathways which recognize Ser/Thr-phosphorylated proteins like those implicated in the mitogen-activated protein kinase (MAPK) pathway.[6,7] To date approx. 300 potential physiological effectors of this family have been identified.[8,9] Currently, only a few small molecules have been shown to stabilize PPIs of some 14-3-3/effector complexes.[10–13] In all these studies the described effectors present the length that is involved in the interaction and not the full sequence. The first reported examples were the fungal toxin Fusicoccin[12] and the plant growth regulator Cotylenin-A,[13] which promote the interaction of 14-3-3 proteins with the human potassium channel Task3[14,15] and the proto-oncogene C-Raf,[13] respectively. Specifically, Fusicoccin showed an EC_{50} value of 2 μM for the stabilization of 14-3-3/Task3 complex, while cotylenin-A is less effective with an EC_{50} of 65 μM for the 14-3-3/C-Raf interaction. More recently, new fusicoccin derivatives (fusicoccanes)[12] and trisubstituted pyrrolidinones (as fusicoccin mimetics)[16,17] have been described as result of different PPIs stabilizer optimization programs. However, their EC_{50} values for stabilization were similar to those of the natural compounds.

Currently more than 100 crystal structures of 14-3-3 proteins with different binding partners and also with their respective inhibitors or stabilizers are available in the PDB[18] (see Table S1 in the ESI†). For example, by analysing the crystal structure of the complex formed by 14-3-3 and C-Raf with cotylenin-A, we observed that the natural product forms two hydrogen bonds with Asp213.[12] In addition, around the central binding channel of 14-3-3 other surface-exposed negatively charged residues are located. Therefore, the design of positively charged compounds that could bind specifically to this anionic region in the vicinity of the central binding channel of 14-3-3 could lead to the modulation (inhibition or stabilization of PPI) of these 14-3-3 interactions with different effectors (Fig. 1).

To identify molecules that could modulate these physiologically highly relevant 14-3-3 PPIs, we performed a screening of cationic supramolecular ligands originating from our homemade library. For the screening we selected ligands which contain the guanidiniocarbonyl pyrrole aka GCP moiety, an arginine mimetic developed by our group.[21] It forms very stable ion pairs with oxyanions such as carboxylates and phosphates significantly more stable than simple guanidinium cations or ammonium ions (as in the natural amino acids arginine or lysine) (Table 1).[19–23] Hence we usually incorporate this binding motif to improve the affinity of peptide-based ligands for proteins. In the search for stabilizers of 14-3-3 protein–protein interactions fluorescence polarization (FP) measurements were carried out, employing 14-3-3ζ/effector systems in which the complex had not formed yet (see Fig. S1, ESI†). The binding partners were fluorescein-labelled, monophosphorylated synthetic

[a] *Institute of Organic Chemistry, University of Duisburg Essen, Universitätstr. 7, 45141, Essen, Germany. E-mail: carsten.schmuck@uni-due.de, christian.ottmann@uni-due.de*
[b] *Bioinformatics and Computational Biophysics, University of Duisburg Essen, Universitätstr. 7, 45141, Essen, Germany. E-mail: daniel.hoffmann@uni-due.de*
[c] *Department of Biomedical Engineering and Institute for Complex Molecular Systems, Technical University Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: c.ottmann@tue.nl*

† Electronic supplementary information (ESI) available: Details of the synthetic procedures and characterization data, as well as details of FP and ITC measurements, and description of the molecular dynamic simulations. See DOI: 10.1039/c8cc07946c

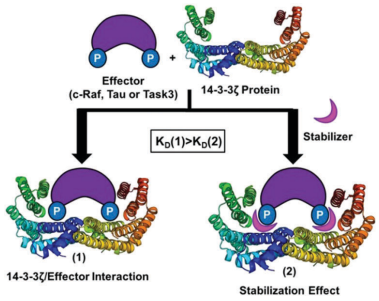
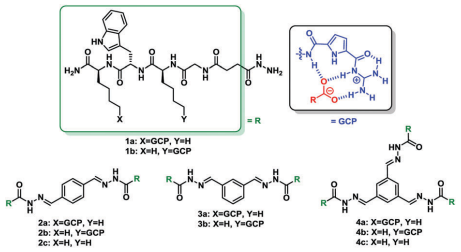


Fig. 1 Stabilizing effect between14-3-3ζ protein (PDB id: 4JDD) and a binding partner (C-Raf, Tau or Task3).

Table 1 EC_{50} values obtained by FP assay of FAM-labelled effectors and 14-3-3ζ titrated with different supramolecular ligands



Cmpds	C-Raf-EC_{50}[a] (μM)	Tau-EC_{50} (μM)	Task3-EC_{50} (μM)
1a	>1000	>1000	>1000
2a	50.0 ± 4.1	89.0 ± 9.9	150.0 ± 7.3
3a	47.0 ± 2.9	>1000	>300
4a	8.4 ± 0.1	1.4 ± 0.2	>200
1b	>1000	>1000	>1000
2b	>1000	22.9 ± 3.9	>300
3b	55.6 ± 0.5	14.5 ± 0.5	147.0 ± 9.0
4b	30.9 ± 2.1	4.7 ± 2.3	>200
2c	>1000	>200	>600
4c	>1000	>200	>1000

[a] 50% effective concentration or concentration of these compounds which is required to stabilize 50% of the interaction between 14-3-3ζ and these effectors.

peptides comprising the phosphorylation sites C-RafpS259 and Task3pS373, as well as the diphosphorylated peptide TaupS214pS324. Binding of these peptides to 14-3-3ζ in the presence and absence of molecules from our library was tested. Several ligands containing the GCP group were identified as new stabilizers of 14-3-3ζ/effector complexes with the best results for C-Raf and Tau binding partners (Table 1).

First, these data highlight the importance of the arginine mimetic in these ligands for the stabilization effect. Compounds which do not have a GCP group, such as **2c** and **4c**, do not show stabilization of the interaction between 14-3-3ζ protein with C-Raf, Tau and Task3; while their corresponding counterparts which have a GCP moiety, such as **2a**, **2b**, **4a** and **4b**, show EC_{50} values in the lower micromolar range. Second, the multivalency of the compounds seems to correlate with their PPI-stabilizing activity. For example, **4a** which has three arms containing one GCP group each increases the stability of 14-3-3ζ/C-Raf complex to a larger extent compared to **3a** or **2a** which only have two arms. In addition, the latter substances show a higher stabilization effect on this interaction than **1a** which possesses a single arm. The same behavior is noted for compounds belonging to series b and also for both series with the 14-3-3ζ partner peptide derived from Tau. Moreover, in terms of selectivity, compound **3a** is noteworthy since it stabilizes specifically the 14-3-3ζ/C-Raf complex with an EC_{50} value of 47.0 ± 2.0 μM without showing an effect on 14-3-3 binding to the peptides derived from Tau and Task3. However, **4a** and **4b** are the most potent compounds of this family, being the first synthetic stabilizers for Tau peptide to 14-3-3ζ and being more potent in comparison with the described Cotylenin A for 14-3-3 ζ/C-Raf interaction, which stabilized the complex with an EC_{50} of around 65 μM.[15] Also the same assay was carried out in absence of 14-3-3ζ protein for the latest compounds, as a control experiment, to ensure that these data are not a response of the binding of these ligands with C-Raf or Tau effectors (see Fig. S3, ESI†). As a result **4a** and **4b** did not bind to C-Raf or Tau.

Table 2 Apparent K_D and the resulting increase in stability (IS) of the 14-3-3ζ/c-Raf and 14-3-3ζ/Tau interactions in the presence of different supramolecular compounds

Cmpds	C-Raf: $K_{D\ app}$ (μM)	IS	Tau: $K_{D\ app}$ (μM)	IS
	16.00 ± 0.60		7.00 ± 0.20	
Control	8.90 ± 0.50[a]	—	3.50 ± 0.10[a]	—
1a	14.40 ± 0.60	×1	5.90 ± 0.10	×1
2a	3.70 ± 0.40	×4	0.56 ± 0.02	×12
3a	7.30 ± 0.20	×2	1.35 ± 0.03	×5
4a	0.19 ± 0.01	×84	0.27 ± 0.01	×26
1b	5.70 ± 0.30[a]	×1.6	2.30 ± 0.10[a]	×1.5
2b	4.50 ± 0.20[a]	×2	0.49 ± 0.01	×14
3b	1.20 ± 0.03	×13	0.37 ± 0.01	×18
4b	0.16 ± 0.01[a]	×56	0.17 ± 0.01[a]	×20

[a] Experiments carried out in a different time and with different batch of protein, effectors and ligands.

Subsequently, we focused on the 14-3-3ζ interactions with the peptide derivates from C-Raf and Tau and determined the apparent K_D values in the absence and presence of our compounds (Table 2). In the presence of 50 μM of these multi-armed compounds we observed an enhancement of the stability of these interactions as reflected in lower apparent K_D values. In line with the EC_{50} data the apparent affinity confirmed the multivalency effect of this family in stabilizing the interactions. Accordingly, compounds with three arms showed a higher stabilization than compounds with two arms, while ligands with one arm were ineffective. **4a**, the most potent compound, enhanced the interaction of 14-3-3ζ with C-Raf and Tau approximately 84-fold and 26-fold, respectively, from a K_D of 16.00 ± 0.60 μM to 0.19 ± 0.01 μM for the binding partner C-Raf and from a K_D of 7.00 ± 0.20 μM to 0.27 ± 0.01 μM for Tau. Similarly, **4b** increased the stability of these interactions, 56-fold in the case of C-Raf and 20-fold for Tau, from a K_D of 8.90 ± 0.50 μM to 0.16 ± 0.01 μM

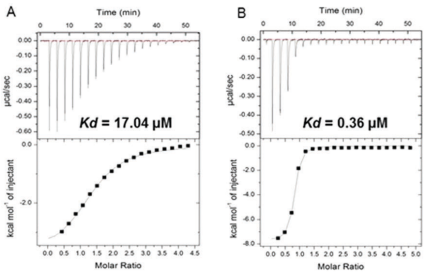


Fig. 2 (A) Isothermal titration calorimetry of binding C-Raf to 14-3-3ζ and (B) the same interaction in the presence of 50 μM of **4b** compound.

for the binding partner C-Raf and from a K_D of 3.50 ± 0.10 μM to 0.17 ± 0.01 μM for Tau.

These observations regarding **4b** were supported by isothermal titration calorimetry (ITC). The observed K_D decreased from 17.04 to 0.36 μM for the complex 14-3-3ζ/C-Raf in the presence of **4b**, enhancing the stabilization 47-fold (Fig. 2). In the case of Tau the K_D decreases from 11.65 to 0.60 μM increasing the stabilization 20-fold (see Fig. S6, ESI†).

Then, we determined by microscale thermophoresis (MST) the binding constants of some of these ligands with 14-3-3ζ (see Fig. S7 and S8, ESI†). The K_D values of **4b**, **4a** and **3a** are 5 μM, 22 μM and 41 μM, respectively. However, **1a** did not show binding to 14-3-3ζ. These results explain why the 3-arm ligands stabilize these interactions better, as their affinity for this protein is higher than its 2-arm counterpart. Conversely, the single arm **1a** does not bind and therefore cannot stabilize. Besides, it is interesting that although the affinity of **4b** is higher than **4a**, **4a** stabilizes the C-Raf interaction more. This could be because its affinity for this complex is higher than in the case of **4b**.

To determine the region of the protein where these compounds most likely bind, molecular dynamics (MD) simulations were carried out using a procedure detailed previously.[27] We studied the complex of 14-3-3ζ/C-Raf (Fig. 3A) in the presence of **4b** (Fig. 3B). However, the flexibility of this molecular system precluded the identification of a single binding pose. Instead, we monitored the volumes explored by the peptidic arms containing the GCP moiety and the central benzene ring in six MD simulations. An exemplary MD simulation is represented in Fig. 3B (see Fig. S9 for all MDs, ESI†). The 3-armed ligand **4b** binds indeed to both the protein (grey surface) and one C-Raf peptide (salmon surface). The central benzene ring of **4b** (blue volume) and one arm (green volume under the blue volume) are located close to the pore of the 14-3-3ζ dimer attaching the ligand to the protein. Another arm of the ligand binds to the C-Raf peptide's C-terminal carboxylic acid (salmon, bottom left corner in the zoomed region), while the third arm is not involved in a salt bridge. The six simulations consistently show the ligand anchoring itself to the protein electrostatically and preventing C-Raf release through steric effects. This is due to the GCP units in ligand **4b** that interact more frequently than Lys units with the anionic Asp/Glu side-chains around the central pore and in regions where the C-Raf peptides also bind (see Fig. S10A, ESI†). The acidic residues most frequently engaged in salt bridges with GCP according to these MD simulations are marked in dark blue in Fig. 4. This binding pattern is fully consistent with the observed stabilization of the 14-3-3/C-Raf interaction by our ligand.

Even though the third arm in **4b** does not form stable salt bridges compared to the other two arms (see Fig. S10B, ESI†),

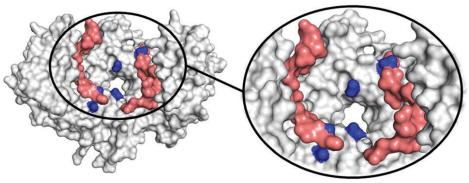


Fig. 4 Anionic amino acids (dark blue) in 14-3-3ζ (grey) and C-Raf peptides frequently involved in salt bridges with the GCP groups (at least 10%) in the MD simulations, averaged over C_2-symmetry equivalent atom pairs in the 14-3-3ζ/C-Raf dimer.

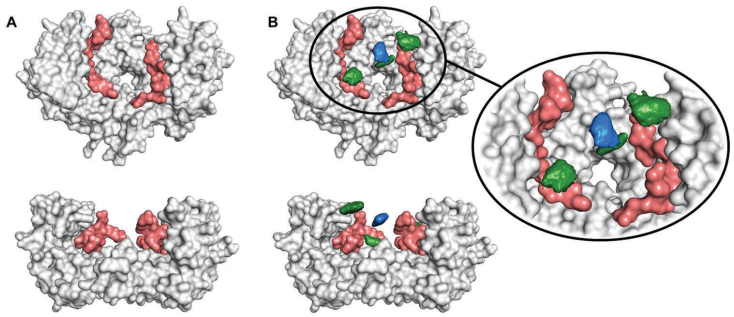


Fig. 3 (A) Two aspects of the complex formed by 14-3-3ζ (grey) and two phosphorylated C-Raf peptides (salmon); (B) MD simulations of 14-3-3ζ dimer (grey) with two phosphorylated C-Raf peptides (salmon), and peptidic arms containing GCP (green) as well as the central benzene ring (blue). Green and blue volumes indicate presence in at least 25% of the MD trajectories.

its presence statistically increases the chances of forming salt bridges during the initial binding, and during unbinding/rebinding events (positive cooperative binding). This could be the reason for the higher affinity of the 3-armed ligand compared to that of the 2-armed ligand.

In summary, we identified a new and potent family of stabilizers of the interaction between 14-3-3ζ and C-Raf or Tau with EC_{50} values in the low micromolar range. Ligands with a higher level of multivalency such as **4a** and **4b** (three arms) are more potent stabilizers for this PPI as shown by both FP measurements and ITC data. The interaction of the protein with the effector peptides is enhanced by nearly two orders of magnitude in the best case. In addition, the area where **4b** most likely is located in the 14-3-3ζ/C-Raf complex was identified *in silico* by Epitopsy, followed by MD to sample the binding events with atomistic resolution, providing a consistent explanation of the observed stabilization. These results should now allow further modifications of this new class of stabilizers to obtain even more potent and selective compounds. A clear advantage of artificial ligands is the much higher synthetic accessibility[23–26] over highly complex natural products which have predominantly been described as PPI stabilizers. To what extent the advantage of natural product PPI stabilizers – cell availability and selectivity – can be achieved with our supramolecular ligands will be the subject of future studies.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 D. C. Fry and L. T. Vassilev, *J. Mol. Med.*, 2005, **83**, 955–963.
2 D. P. Ryan and J. M. Matthews, *Curr. Opin. Struct. Biol.*, 2005, **15**, 441–466.
3 J.-C. Carry and C. Garcia-Echeverria, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 2480–2485.
4 G. Fischer, M. Rossmann and M. Hyvönen, *Curr. Opin. Biotechnol.*, 2015, **35**, 78–85.
5 M. Aeluri, S. Chamakuri, B. Dasari, S. K. R. Guduru, R. Jimmidi, S. Jogula and P. Arya, *Chem. Rev.*, 2014, **114**, 4640–4694.
6 D. Berg, C. Holzmann and O. Riess, *Nat. Rev. Neurosci.*, 2003, **4**, 752–762.
7 X. Yang, W. H. Lee, F. Sobott, E. Papagrigoriou, C. V. Robinson, J. G. Grossmann, M. Sundström, D. A. Doyle and J. M. Elkins, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 17237–17242.
8 G. P. H. van Heusden, *IUBMB Life*, 2005, **57**, 623–629.
9 C. Johnson, S. Crowther, M. J. Stafford, D. G. Campbell, R. Toth and C. MacKintosh, *Biochem. J.*, 2010, **427**, 69–78.
10 A. Kaplan, B. Morquette, A. Kroner, S. Leong, C. Madwar, R. Sanz, S. L. Banerjee, J. Antel, N. Bisson, S. David and A. E. Fournier, *Neuron*, 2017, **93**, 1082–1093.
11 M. Mori, G. Vignaroli and M. Botta, *Drug Discovery Today: Technol.*, 2013, **10**, e541–e547.
12 R. G. Doveston, A. Kuusk, S. A. Andrei, S. Leysen, Q. Cao, M. P. Castaldi, A. Hendricks, L. Brunsveld, H. Chen, H. Boyd and C. Ottmann, *FEBS Lett.*, 2017, **591**, 2449–2457.
13 L. M. Stevers, E. Sijbesma, M. Botta, C. MacKintosh, T. Obsil, I. Landrieu, Y. Cau, A. J. Wilson, A. Karawajczyk, J. Eickhoff, J. Davis, M. Hann, G. O'Mahony, R. G. Doveston, L. Brunsveld and C. Ottmann, *J. Med. Chem.*, 2018, **61**, 3755–3778.
14 C. Anders, Y. Higuchi, K. Koschinsky, M. Bartel, B. Schumacher, P. Thiel, H. Nitta, R. Preisig-Müller, G. Schlichthörl, V. Renigunta, J. Ohkanda, J. Daut, N. Kato and C. Ottmann, *Chem. Biol.*, 2013, **20**, 583–593.
15 M. Molzan, S. Kasper, L. Röglin, M. Skwarczynska, T. Sassa, T. Inoue, F. Breitenbuecher, J. Ohkanda, N. Kato, M. Schuler and C. Ottmann, *ACS Chem. Biol.*, 2013, **8**, 1869–1875.
16 S. Bittner, T. Budde, H. Wiendl and S. G. Meuth, *Brain Pathol.*, 2010, **20**, 999–1009.
17 M. Kilisch, O. Lytovchenko, B. Schwappach, V. Renigunta and J. Daut, *Pflügers Arch. Eur. J. Physiol.*, 2015, **467**, 1105–1120.
18 R. Rose, S. Erdmann, S. Bovens, A. Wolf, M. Rose, S. Hennig, H. Waldmann and C. Ottmann, *Angew. Chem.*, 2010, **49**, 4129–4132.
19 A. Richter, R. Rose, C. Hedberg, H. Waldmann and C. Ottmann, *Chem*, 2012, **18**, 6520–6527.
20 PDB: 5J31, 5EWZ, 5EXA, 5D2D, 5D3F, 4ZDR, 4WRQ, 4N7G, 4N7Y, 4N84, 4IEA, 4IHL... (see ESI†).
21 M. Ehlers, J.-N. Grad, S. Mittal, D. Bier, M. Mertel, L. Ohl, M. Bartel, J. Briels, M. Heimann, C. Ottmann, E. Sanchez-Garcia, D. Hoffmann and C. Schmuck, *ChemBioChem*, 2017, **19**, 591–595.
22 C. Schmuck, *Coord. Chem. Rev.*, 2006, **250**, 3053–3067.
23 Q. Q. Jiang, L. Bartsch, W. Sicking, P. R. Wich, D. Heider, D. Hoffmann and C. Schmuck, *Org. Biomol. Chem.*, 2013, **11**, 1631–1639.
24 Q. Q. Jiang, W. Sicking, M. Ehlers and C. Schmuck, *Chem. Sci.*, 2015, **6**, 1792–1800.
25 M. Li, M. Ehlers, S. Schlesiger, E. Zellermann, S. K. Knauer and C. Schmuck, *Angew. Chem.*, 2016, **55**, 598–601.
26 M. Li, S. Schlesiger, S. K. Knauer and C. Schmuck, *Angew. Chem.*, 2015, **54**, 2941–2944.
27 J.-N. Grad, A. Gigante, C. Wilms, J. N. Dybowski, L. Ohl, C. Ottmann, C. Schmuck and D. Hoffmann, *J. Chem. Inf. Model.*, 2018, **58**, 315–327.

A New Class of Supramolecular Ligands Stabilizes 14-3-3 Protein-Protein Interactions by up to Two Orders of Magnitude

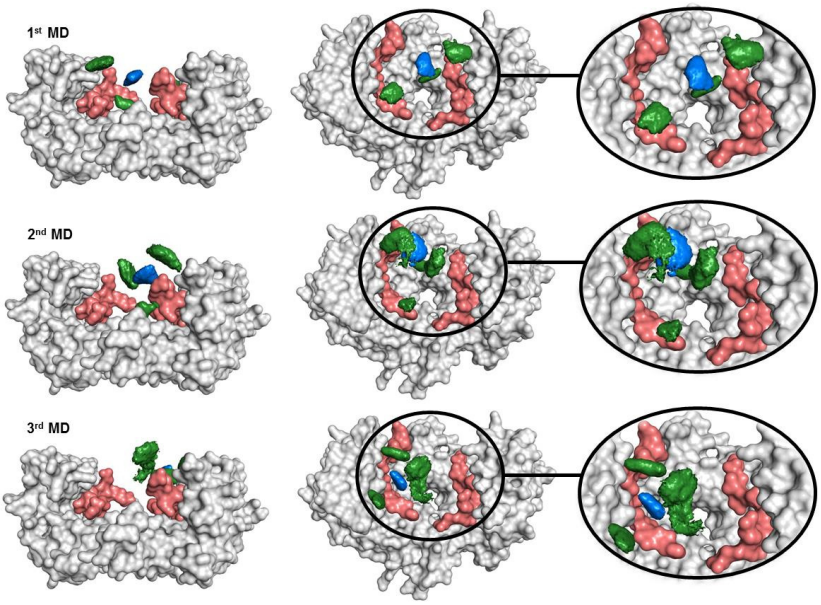
A. Gigante, J.-N. Grad, J. Briels, M. Bartel, D. Hoffmann, C. Ottmann, and C. Schmuck

Supporting Information

[Pages S2 to S8 not reproduced]

5. Molecular Dynamics simulations (MDs)

MD simulations were carried out with Gromacs 4.6.7[1] using the methodology and parameters described in the literature.[2] 12 simulations were produced: a set of 6 pilot simulations of 30 ns to sample the binding event (Figures 3, 4, S8), followed by a set of 6 production simulations of 250 ns, amounting to 1.5 μs, for the statistical analysis of salt bridge formation (Figure S9). We determined the frequency of salt bridge formation between ligand **4b** and the 14-3-3/C-Raf receptor by measuring the N···O distance between their charged groups (Lys ε-ammonium group, GCP guanidinium group, Asp/Glu side-chain carboxylate groups, C-Raf peptide *C*-terminal carboxylate group) using a cutoff value of 3.2 Å. Figure S9A shows how often the 6 basic residues in ligand **4b** interacted with acidic groups from the 14-3-3ζ/C-Raf complex. Overall GCP formed salt bridges more frequently than Lys. Figure S9B shows that in MD 2, 3, and 6 the third arm did not actively participate in salt bridge formation.



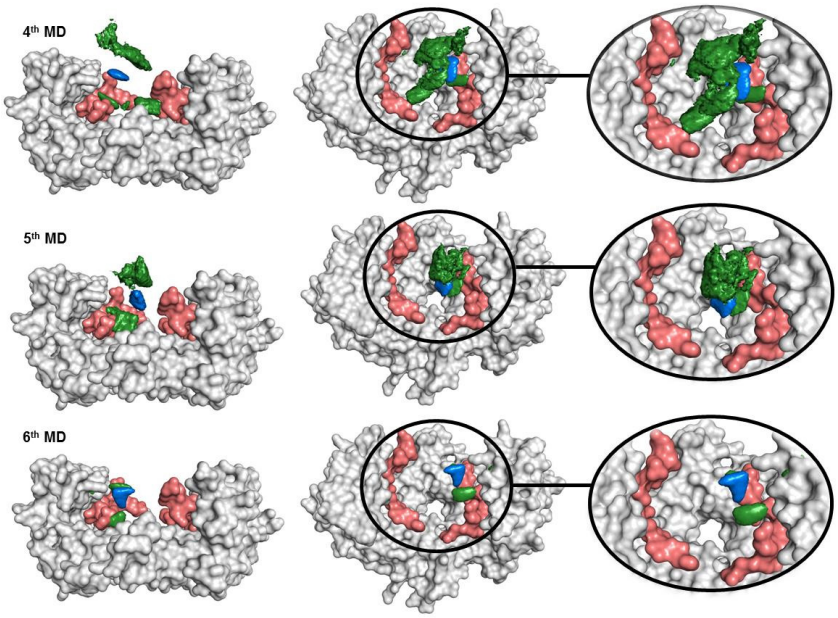


Figure S8. Six MD simulations of 14-3-3ζ dimer (grey) with C-Raf peptides (salmon) and peptidic arms containing GCP (green) as well as the central benzene ring (blue). Green and blue volumes indicate presence in at least 25% of the MD trajectories.

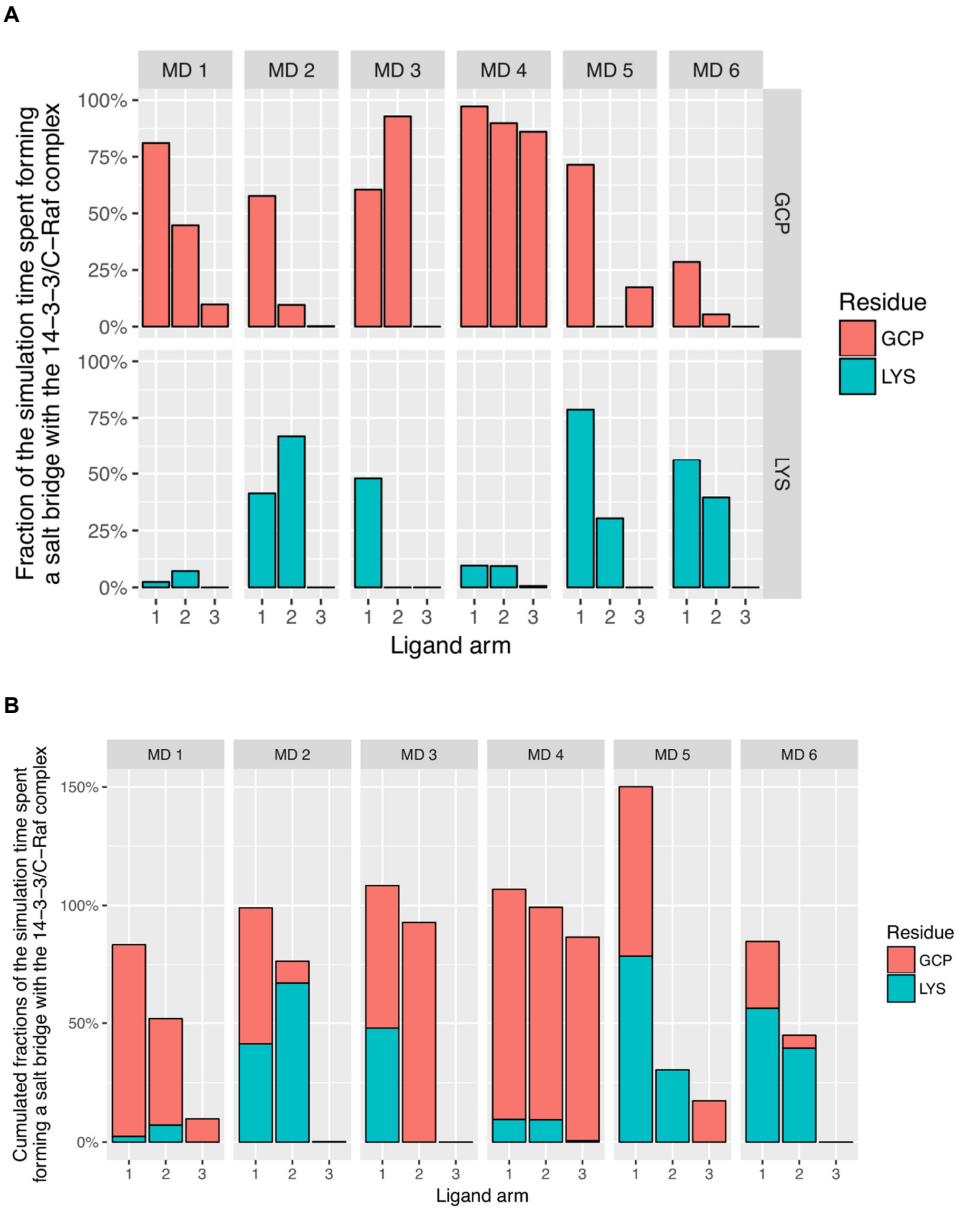


Figure S9. Salt bridge frequency in the MD simulations involving the 14-3-3/C-Raf/**4b** complex. (A) The three chemically equivalent arms in ligand **4b** are labeled 1, 2, and 3 by decreasing frequency of binding. (B) combines the same results from (A) in one graphic to highlight the relative contribution of each arm.

[Pages S12 to S34 not reproduced]

3.2 Discussion

3.2.1 Molecular modeling

3.2.1.1 14-3-3:c-Raf complex

14-3-3 isoforms are evolutionarily highly conserved (Wang and Shakes 1996) and adopt a characteristic U-shaped quaternary structure made of nine α-helices stacked anti-parallel to each other. 14-3-3 forms complexes with client proteins via a phosphopeptide-binding site (residues Lys^{49}, Arg^{56}, Arg^{127} and Tyr^{128}) located in an amphipathic groove (Liu et al. 1995; Yaffe et al. 1997). c-Raf becomes a client of 14-3-3 upon phosphorylation of residues Ser^{233}, Ser^{259} or Ser^{621} (Luo et al. 1995; Tzivion et al. 1998; Yip-Schneider et al. 2000). Although the full tertiary structure of c-Raf is unknown (Hatzivassiliou et al. 2010), it is assumed that Ser^{233} and Ser^{259} can bind simultaneously to both binding sites of a 14-3-3 dimer to form a bidentate complex. This assumption is based on a co-crystal of 14-3-3ζ with a di-phosphorylated $pSer^{233}$, $pSer^{259}$ c-Raf fragment (Gln^{229} to His^{264}) and on the cooperative effect observed in binding studies with di-phosphorylated peptides (Molzan and Ottmann 2012).

The 14-3-3 *C*-termini are poorly resolved in crystal structures (Williams et al. 2011). Since they interact weakly with the amphipathic grooves and get easily displaced by phosphopeptides (Silhan et al. 2004), they are often truncated in MD simulations, in an effort to reduce the box size and avoid poor sampling statistics. One notable exception is a Hamiltonian replica exchange investigation of full-length 14-3-3 dimers and monomers binding to phosphopeptides for 6.7 µs of enhanced-sampling MD simulation time (Nagy et al. 2017). The study confirmed the poor sampling of the 14-3-3 *C*-termini, and despite their interference with phosphopeptides binding pathways in a few simulations, the same experiments carried out with truncated 14-3-3 *C*-termini resulted in reasonable agreement with experimental data.

A truncated version of the 14-3-3ζ dimer ("14-3-3ΔC", sequence Met^{1} to Ser^{230}) was used for this work, with two identical c-Raf fragments phosphorylated at Ser^{259} ("c-Raf(pS^{259})", sequence Arg^{254} to His^{264}) and initially located in the two 14-3-3ΔC binding sites. Using the same c-Raf sequence in both binding sites helped preserving the C_2-symmetry of the receptor and allowed for a more direct comparison to the results of the fluorescence polarization assay, where a similarly short fluorescent probe (5-FAM)–RQRSTpSTPNVH–COO^- was used. A model of the 14-3-3ΔC:c-Raf(pS^{259}) dimer was built using PDB entry 4ihl (Molzan et al. 2013).

3.2.1.2 QQJ-096 stabilizer

The QQJ-096 ligand is composed of three peptidic arms with sequence GKWK connected to an aromatic tri-hydrazone *via* a succinic acid linker (Figure 5 in Work 1). On each arm, one lysine is decorated with a guanidiniocarbonyl pyrrole (GCP) cation. This scaffold was originally investigated in a combinatoric library of multi-armed peptidic inhibitors of β-tryptase (Jiang et al. 2013), where the substitution of arginine by the artificial mimetic GCP increased the binding affinity by two orders of magnitude. The Lys–Trp–Lys motif was explored in a subsequent fluorescence study (Wang et al. 2014) and retained for a new generation of multi-armed peptidic inhibitors of β-tryptase featuring a central aryl ring (Jiang et al. 2015). This class of inhibitors was later discovered to stabilize the 14-3-3:c-Raf interaction in the

sub-micromolar range, prompting further research to develop a 14-3-3:c-Raf–tailored stabilizer in the nanomolar range (Bartel 2015, chapter 5).

The parameterization procedure is detailed in section "Molecular Dynamics Simulations" in Work 1. The ligand was assembled in TLEaP and equilibrated in explicit solvent with NaCl 0.15 mol l^{-1} for several nanoseconds.

3.2.2 MD simulations and data analysis

Given the size of the molecular system, optimal MD initial conditions were needed to reduce the sampling time. EGs calculated with two probes extracted from the stabilizer revealed a probable binding site above the 14-3-3 pore. Using this prediction, the stabilizer was initially placed above the pore and 10 Å away from the c-Raf fragments, in an effort to gain some insight into the binding pathways that lead to the amphipathic grooves. This precaution proved unfruitful, as the stabilizer failed to make contact with the c-Raf fragments within the first 50 ns in half of the simulations (Figure S1 in Work 1). A new series of simulations were carried out with the stabilizer much closer to the c-Raf fragments, at which point contact could be observed within 2 ns. This series was sampled for 6×250 ns.

Due to the highly dynamic nature of the binding, it was not possible to extract meaningful binding poses for the stabilizer. Instead, MD occupancies were calculated for the GCP and central aryl moieties (Figures 3 in Work 2 and S8 in Work 2), in order to visualize the spatial extent of the three arms. Most of the 6 trajectories have well-localized blobs of density, which give a qualitative understanding of the interaction specificity.

For a more quantitative approach, salt bridge formation between QQJ-096 and the receptor was monitored in the trajectories. Plotting the frequency of salt bridges from the point of view of acidic residues in the receptor revealed Glu^{14} and Glu^{17} to be key partners of QQJ-096 (Figure 3.1), with a minor contribution from Asp^{223}. These statistics were mapped to the protein surface for better visualization (Figure 4 in Work 2). The two glutamates are located next to each other in space (7 Å), and relatively close to their C_2-related pair in the second chain of the 14-3-3ΔC dimer (15 Å between the two Glu^{17}). The proximity of these acidic residues coupled to the scarcity of solvent-exposed basic residues in the immediate vicinity[6] leads to the formation of a negatively charged binding site above the pore, ideally suited to a polycationic ligand.

Plotting the salt bridge frequency from the point of view of cationic groups in QQJ-096 revealed that at least one GCP and one Lys did not engage in salt bridges (Figure S9 in Work 2), albeit not necessarily from the same arm (Figure S9 in Work 2, panel "MD 5"). This observation is consistent with experimentally determined EC_{50} values (Table 1 in Work 2), which shows a minor decrease in binding affinity between QQJ-096 (30.9 µM ± 2.1 µM, compound 4b) and its two-arm equivalent (55.6 µM ± 0.5 µM, compound 3b).

Considering the large volume of the 14-3-3 cleft, it would be interesting to determine the stoichiometry of the binding. The previous series of MDs handled the case of a 2:2:1 ratio of the 14-3-3ΔC:c-Raf(pS^{259}):QQJ-096 complex. To handle the 2:2:2 case, the last frame of each of the six MDs was used as the starting condition for a new MD, with an additional QQJ-096 molecule manually placed so as to form at least two salt bridges with 14-3-3ΔC. The rationale behind

[6] Basic residues inside the pore are physically inaccessible to QQJ-096, residues Lys^{49}, Arg^{56}, Arg^{60}, Arg^{127} from the amphipathic groove are occluded by the c-Raf(pS^{259}) fragments, leaving only Lys^{11} in the cleft.

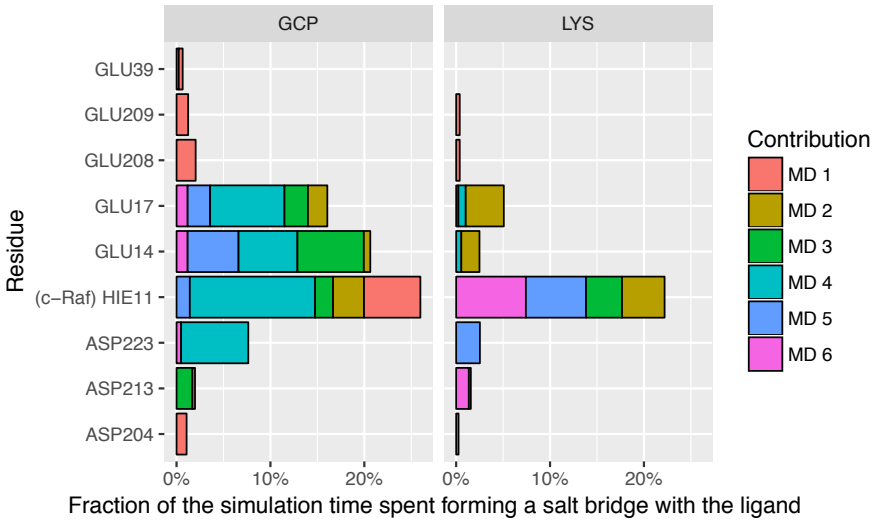


Figure 3.1: Frequency of binding of acidic amino acids (Glu, Asp, *C*-terminal carboxylate of c-Raf(pS^{259})) with the positively charged groups in QQJ-096. Only the most frequent binding partners are represented. The frequency is averaged across the C_2-related residues in the receptor and the C_3-related residues in QQJ-096.

this approach was the concern that the high concentration of positive charges on QQJ-096 would cause the new, randomly placed stabilizer to be electrostatically driven out of the cleft by the original, optimally-anchored stabilizer.

The 2:2:2 stoichiometry was sampled for 6×50 ns. In 3 cases out of 6, the extra QQJ-096 partially left the cleft due to steric occlusion of the binding site by the original stabilizer. Plotting the frequency of salt bridges from the point of view of the cationic groups in QQJ-096 showed overall that two GCPs in the first QQJ-096 and one GCP in the second QQJ-096 were involved in salt bridges with the receptor (Figure 3.2). Although the cleft is in principle sufficiently spacious to accommodate for two QQJ-096 ligands, the plot suggested a non-additive effect, probably due to saturation of the binding site. A similar conclusion was reached in MD simulations with a 4-armed analogue of QQJ-096 (Bartsch et al. 2019).

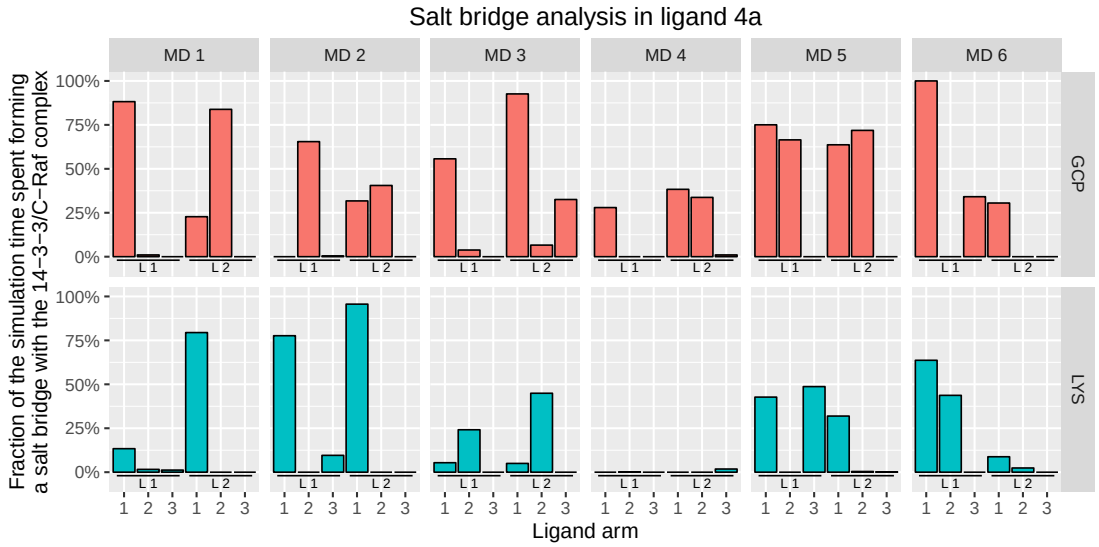


Figure 3.2: Salt bridge frequency for the GCP and Lys groups in QQJ-096 (same as Figure S9 in Work 2, but with 2 equivalents of QQJ-096, named L1 and L2). The bars represent the fraction of the simulation time spent in a salt bridge between any basic residue in QQJ-096 and any acidic residues in the 14-3-3ΔC:c-Raf(pS^{259}) receptor. A salt bridge is formed when the distance between a nitrogen atom in the hydrogen-bond donor (Lys ε-ammonium group, GCP guanidinium group) and an oxygen atom in the hydrogen-bond receptor (Asp and Glu side-chain carboxylate groups, c-Raf(pS^{259}) peptide *C*-terminal carboxylate group) is smaller than 3.2 Å.

Bibliography

van Aalten DM, Bywater R, Findlay JB, Hendlich M, Hooft RW, Vriend G (1996) PRODRG, a Program for Generating Molecular Topologies and Unique Molecular Descriptors From Coordinates of Small Molecules. In: Journal of Computer-Aided Molecular Design 10(3), pp. 255–262. doi:10.1007/BF00355047

Aghazadeh Y, Papadopoulos V (2016) The Role of the 14-3-3 Protein Family in Health, Disease, and Drug Development. In: Drug Discovery Today 21(2), pp. 278–287. doi:10.1016/j.drudis.2015.09.012

Aitken A (2002) Functional Specificity in 14-3-3 Isoform Interactions Through Dimer Formation and Phosphorylation. Chromosome Location of Mammalian Isoforms and Variants. In: Plant Molecular Biology 50(6), pp. 993–1010. doi:10.1023/A:1021261931561

Allen FH (2002) The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. In: Acta Crystallographica. Section B, Structural Science 58(3 Part 1), pp. 380–388. doi:10.1107/s0108768102003890

Anderson AC (2003) The Process of Structure-Based Drug Design. In: Chemistry & Biology (Oxford, United Kingdom) 10(9), pp. 787–797. doi:10.1016/j.chembiol.2003.09.002

Anzt H, Bach F, Druskat S, Löffler F, Loewe A, Renard BY, Seemann G, Struck A, Achhammer E, Aggarwal P, Appel F, Bader M, Brusch L, Busse C, Chourdakis G, Dabrowski PW, Ebert P, Flemisch B, Friedl S, Fritzsch B, Funk MD, Gast V, Goth F, Grad JN, Hegewald J, Hermann S, Hohmann F, Janosch S, Kutra D, Linxweiler J, Muth T, Peters-Kottig W, Rack F, Raters FHC, Rave S, Reina G, Reißig M, Ropinski T, Schaarschmidt J, Seibold H, Thiele JP, Uekermann B, Unger S, Weeber R (2021) An Environment for Sustainable Research Software in Germany and Beyond: Current State, Open Challenges, and Call for Action [version 2; peer review: 2 approved]. In: F1000Research 9(295). doi:10.12688/f1000research.23224.2

Ascierto PA, Kirkwood JM, Grob JJ, Simeone E, Grimaldi AM, Maio M, Palmieri G, Testori A, Marincola FM, Mozzillo N (2012) The Role of BRAF V600 Mutation in Melanoma. In: Journal of Translational Medicine 10, p. 85. doi:10.1186/1479-5876-10-85

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. In: Proceedings of the National Academy of Sciences of the United States of America 98(18), pp. 10037–10041. doi:10.1073/pnas.181342398

Barlow DJ, Thornton JM (1983) Ion-Pairs in Proteins. In: Journal of Molecular Biology 168(4), pp. 867–885. doi:10.1016/s0022-2836(83)80079-5

Bartel M (2015) Molecular Concepts for the Modulation of 14-3-3 Protein-Protein Interactions: Small-Molecule Stabilization, Supramolecular Multivalent Binders and Peptidic Inhibition. PhD thesis. Department of Biomedical Engineering, University of Technology, Eindhoven. URL: https://pure.tue.nl/ws/files/24302433/20160609_Bartel.pdf

Bartel M, Schäfer A, Stevers LM, Ottmann C (2014) Small Molecules, Peptides and Natural Products: Getting a Grip on 14-3-3 Protein–Protein Modulation. In: Future Medicinal Chemistry 6(8), pp. 903–921. doi:10.4155/fmc.14.47

Bartsch L, Bartel M, Gigante A, Iglesias-Fernández J, Ruiz-Blanco YB, Beuck C, Briels J, Toetsch N, Bayer P, Sanchez-Garcia E, Ottmann C, Schmuck C (2019) Multivalent Ligands With Tailor-Made Anion Binding Motif as Stabilizers of Protein–Protein Interactions. In: ChemBioChem: A European Journal of Chemical Biology 20(23), pp. 2921–2926. doi:10.1002/cbic.201900288

Basma M, Sundara S, Calgan D, Vernali T, Woods RJ (2001) Solvated Ensemble Averaging in the Calculation of Partial Atomic Charges. In: Journal of Computational Chemistry 22(11), pp. 1125–1137. doi:10.1002/jcc.1072

Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. In: Journal of Physical Chemistry 97(40), pp. 10269–10280. doi:10.1021/j100142a004

Behnel S, Bradshaw R, Citro C, Dalcin L, Seljebotn DS, Smith K (2011) Cython: The Best of Both Worlds. In: Computing in Science & Engineering 13(2), pp. 31–39. doi:10.1109/MCSE.2010.118

Berendsen HJC, Grigera JR, Straatsma TP (1987) The Missing Term in Effective Pair Potentials. In: The Journal of Physical Chemistry 91(24), pp. 6269–6271. doi:10.1021/j100308a038

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. In: Nucleic Acids Research 28(1), pp. 235–242. doi:10.1093/nar/28.1.235

Bernardi RC, Melo MCR, Schulten K (2015) Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. In: Biochimica et Biophysica Acta 1850(5), pp. 872–877. doi:10.1016/j.bbagen.2014.10.019

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. In: Journal of Molecular Biology 112(3), pp. 535–542. doi:10.1016/s0022-2836(77)80200-3

Besler BH, Merz KM, Kollman PA (1990) Atomic Charges Derived From Semiempirical Methods. In: Journal of Computational Chemistry 11(4), pp. 431–439. doi:10.1002/jcc.540110404

Betz RM, Walker RC (2015) Paramfit: Automated Optimization of Force Field Parameters for Molecular Dynamics Simulations. In: Journal of Computational Chemistry 36(2), pp. 79–87. doi:10.1002/jcc.23775

Biedermann F, Schneider HJ (2016) Experimental Binding Energies in Supramolecular Complexes. In: Chemical Reviews 116(9), pp. 5216–5300. doi:10.1021/acs.chemrev.5b00583

Bier D, Rose R, Bravo-Rodriguez K, Bartel M, Ramirez-Anguita JM, Dutt S, Wilch C, Klärner FG, Sanchez-Garcia E, Schrader T, Ottmann C (2013) Molecular Tweezers

Modulate 14-3-3 Protein–Protein Interactions. In: Nature Chemistry 5(3), pp. 234–239. doi:10.1038/nchem.1570

Bikadi Z, Hazai E (2009) Application of the PM6 Semi-Empirical Method to Modeling Proteins Enhances Docking Accuracy of AutoDock. In: Journal of Cheminformatics 1, p. 15. doi:10.1186/1758-2946-1-15

Blasco RB, Francoz S, Santamaría D, Cañamero M, Dubus P, Charron J, Baccarini M, Barbacid M (2011) c-Raf, but Not B-Raf, Is Essential for Development of K-Ras Oncogene-Driven Non-Small Cell Lung Carcinoma. In: Cancer Cell 19(5), pp. 652–663. doi:10.1016/j.ccr.2011.04.002

Bodnarchuk MS (2016) Water, Water, Everywhere… It's Time to Stop and Think. In: Drug Discovery Today 21(7), pp. 1139–1146. doi:10.1016/j.drudis.2016.05.009

Bonn-Breach R, Gu Y, Jenkins J, Fasan R, Wedekind J (2019) Structure of Sonic Hedgehog Protein in Complex With Zinc(II) and Magnesium(II) Reveals Ion-Coordination Plasticity Relevant to Peptide Drug Design. In: Acta Crystallographica. Section D, Structural Biology 75(Pt 11), pp. 969–979. doi:10.1107/S2059798319012890

Borek D, Minor W, Otwinowski Z (2003) Measurement Errors and Their Consequences in Protein Crystallography. In: Acta Crystallographica. Section D, Biological Crystallography 59(Pt 11), pp. 2031–2038. doi:10.1107/s0907444903020924

Borukhov I, Andelman D, Orland H (1997) Steric Effects in Electrolytes: A Modified Poisson-Boltzmann Equation. In: Physical Review Letters 79(3), pp. 435–438. doi:10.1103/PhysRevLett.79.435

Brenke R, Kozakov D, Chuang GY, Beglov D, Hall D, Landon MR, Mattos C, Vajda S (2009) Fragment-Based Identification of Druggable 'Hot Spots' of Proteins Using Fourier Domain Correlation Techniques. In: Bioinformatics (Oxford, England) 25(5), pp. 621–627. doi:10.1093/bioinformatics/btp036

Bumcrot DA, Takada R, McMahon AP (1995) Proteolytic Processing Yields Two Secreted Forms of Sonic Hedgehog. In: Molecular and Cellular Biology 15(4), pp. 2294–2303. doi:10.1128/MCB.15.4.2294

Callenberg KM, Choudhary OP, de Forest GL, Gohara DW, Baker NA, Grabe M (2010) APBSmem: A Graphical Interface for Electrostatic Calculations at the Membrane. In: PLoS One 5(9). doi:10.1371/journal.pone.0012722

Cardin AD, Weintraub HJ (1989) Molecular Modeling of Protein-Glycosaminoglycan Interactions. In: Arteriosclerosis (Dallas) 9(1), pp. 21–32. doi:10.1161/01.atv.9.1.21

Carugo O, Bordo D (1999) How Many Water Molecules Can Be Detected by Protein Crystallography? In: Acta Crystallographica. Section D, Biological Crystallography 55(Pt 2), pp. 479–483. doi:10.1107/s0907444998012086

Di Cera E (1998) Site-Specific Thermodynamics: Understanding Cooperativity in Molecular Recognition. In: Chemical Reviews 98(4), pp. 1563–1592. doi:10.1021/cr960135g

Di Cera E (2020) Mechanisms of Ligand Binding. In: Biophysics Reviews 1(1), p. 011303. doi:10.1063/5.0020997

Černý V (1985) Thermodynamical Approach to the Traveling Salesman Problem: An Efficient Simulation Algorithm. In: Journal of Optimization Theory and Applications 45(1), pp. 41–51. doi:10.1007/BF00940812

Chakraborty P, Di Cera E (2017) Induced Fit Is a Special Case of Conformational Selection. In: Biochemistry 56(22), pp. 2853–2859. doi:10.1021/acs.biochem.7b00340

Chen R, Weng Z (2002) Docking Unbound Proteins Using Shape Complementarity, Desolvation, and Electrostatics. In: Proteins 47(3), pp. 281–294. doi:10.1002/prot.10092

Chirlian LE, Francl MM (1987) Atomic Charges Derived From Electrostatic Potentials: A Detailed Study. In: Journal of Computational Chemistry 8(6), pp. 894–905. doi:10.1002/jcc.540080616

Chu VB, Bai Y, Lipfert J, Herschlag D, Doniach S (2007) Evaluation of Ion Binding to DNA Duplexes Using a Size-Modified Poisson-Boltzmann Theory. In: Biophysical Journal 93(9), pp. 3202–3209. doi:10.1529/biophysj.106.099168

Chu VB, Bai Y, Lipfert J, Herschlag D, Doniach S (2008) A Repulsive Field: Advances in the Electrostatics of the Ion Atmosphere. In: Current Opinion in Chemical Biology 12(6), pp. 619–625. doi:10.1016/j.cbpa.2008.10.010

Chukkapalli G, Karpik SR, Ethier CR (1999) A Scheme for Generating Unstructured Grids on Spheres With Application to Parallel Computation. In: Journal of Computational Physics 149(1), pp. 114–127. doi:10.1006/jcph.1998.6146

Cieplak P, Cornell WD, Bayly C, Kollman PA (1995) Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. In: Journal of Computational Chemistry 16(11), pp. 1357–1377. doi:10.1002/jcc.540161106

Clapham DE (2007) Calcium Signaling. In: Cell 131(6), pp. 1047–1058. doi:10.1016/j.cell.2007.11.028

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL (2009) Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. In: Bioinformatics (Oxford, England) 25(11), pp. 1422–1423. doi:10.1093/bioinformatics/btp163

Collette A (2010) ANFFT. Google Code. URL: https://code.google.com/archive/p/anfft/

Cooley JW, Tukey JW (1965) An Algorithm for the Machine Calculation of Complex Fourier Series. In: Mathematics of Computation 19(90), pp. 297–301. doi:10.1090/S0025-5718-1965-0178586-1

Cooper DR, Boczek T, Grelewska K, Pinkowska M, Sikorska M, Zawadzki M, Derewenda Z (2007) Protein Crystallization by Surface Entropy Reduction: Optimization of the SER Strategy. In: Acta Crystallographica. Section D, Biological Crystallography 63(Pt 5), pp. 636–645. doi:10.1107/S0907444907010931

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. In: Journal of the American Chemical Society 117(19), pp. 5179–5197. doi:10.1021/ja00124a002

Cornell WD, Cieplak P, Bayly CI, Kollmann PA (1993) Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. In: Journal of the American Chemical Society 115(21), pp. 9620–9631. doi:10.1021/ja00074a030

Cox SR, Williams DE (1981) Representation of the Molecular Electrostatic Potential by a Net Atomic Charge Model. In: Journal of Computational Chemistry 2(3), pp. 304–323. doi:10.1002/jcc.540020312

Cramer F (2007) Emil Fischer's Lock-and-Key Hypothesis After 100 Years—Towards a Supracellular Chemistry. In: Perspectives in Supramolecular Chemistry. Ed. by Behr JP. Wiley-Blackwell. Chap. 1, pp. 1–23. ISBN:9780470511411. doi:10.1002/9780470511411.ch1

Cramer F, Freist W (1987) Molecular Recognition by Energy Dissipation, a New Enzymatic Principle: The Example Isoleucine-Valine. In: Accounts of Chemical Research 20(3), pp. 79–84. doi:10.1021/ar00135a001

Csermely P, Palotai R, Nussinov R (2010) Induced Fit, Conformational Selection and Independent Dynamic Segments: An Extended View of Binding Events. In: Trends in Biochemical Sciences 35(10), pp. 539–546. doi:10.1016/j.tibs.2010.04.009

Cui J, Song Y, Han X, Hu J, Chen Y, Chen X, Xu X, Xing Y, Lu H, Cai L (2020) Targeting 14-3-3ζ Overcomes Resistance to Epidermal Growth Factor Receptor-Tyrosine Kinase Inhibitors in Lung Adenocarcinoma via BMP2/Smad/ID1 Signaling. In: Frontiers in Oncology 10, p. 542007. doi:10.3389/fonc.2020.542007

Cullen MJP (1974) Integrations of the Primitive Equations on a Sphere Using the Finite Element Method. In: Quarterly Journal of the Royal Meteorological Society 100(426), pp. 555–562. doi:10.1002/qj.49710042605

Cullen MJP, Hall CD (1979) Forecasting and General Circulation Results From Finite Element Models. In: Quarterly Journal of the Royal Meteorological Society 105(445), pp. 571–592. doi:10.1002/qj.49710544506

Daya-Grosjean L, Couvé-Privat S (2005) Sonic Hedgehog Signaling in Basal Cell Carcinomas. In: Cancer Letters 225(2), pp. 181–192. doi:10.1016/j.canlet.2004.10.003

Daye LR, Gibson W, Williams KP (2010) Development of a High Throughput Screening Assay for Inhibitors of Hedgehog-Heparin Interactions. In: International Journal of High Throughput Screening 1, pp. 69–80. doi:10.2147/IJHTS.S7704

Dennis S, Kortvelyesi T, Vajda S (2002) Computational Mapping Identifies the Binding Sites of Organic Solvents on Proteins. In: Proceedings of the National Academy of Sciences of the United States of America 99(7), pp. 4290–4295. doi:10.1073/pnas.062398499

Derewenda ZS (2010) Application of Protein Engineering to Enhance Crystallizability and Improve Crystal Properties. In: Acta Crystallographica. Section D, Biological Crystallography 66(Pt 5), pp. 604–615. doi:10.1107/S090744491000644X

Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. In: Journal of the American Chemical Society 107(13), pp. 3902–3909. doi:10.1021/ja00299a024

Dickson CJ, Madej BD, Skjevik AA, Betz RM, Teigen K, Gould IR, Walker RC (2014) Lipid14: The Amber Lipid Force Field. In: Journal of Chemical Theory and Computation 10(2), pp. 865–879. doi:10.1021/ct4010307

Ditchfield R, Hehre WJ, Pople JA (1971) Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. In: The Journal of Chemical Physics 54(2), pp. 724–728. doi:10.1063/1.1674902

Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA (2007) PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. In: Nucleic Acids Research 35(Web Server issue), W522–W525. doi:10.1093/nar/gkm276

Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: An Automated Pipeline for the Setup of Poisson–Boltzmann Electrostatics Calculations. In: Nucleic Acids Research 32(Web Server issue), W665–W667. doi:10.1093/nar/gkh381

Donald JE, Kulp DW, DeGrado WF (2011) Salt Bridges: Geometrically Specific, Designable Interactions. In: Proteins 79(3), pp. 898–915. doi:10.1002/prot.22927

Edmundson JR (1992) The Distribution of Point Charges on the Surface of a Sphere. In: Acta Crystallographica Section A 48(1), pp. 60–69. doi:10.1107/S0108767391009133

Ehlers M, Grad JN, Mittal S, Bier D, Mertel M, Ohl L, Bartel M, Briels J, Heimann M, Ottmann C, Sanchez-Garcia E, Hoffmann D, Schmuck C (2018) Rational Design, Binding Studies, and Crystal-Structure Evaluation of the First Ligand Targeting the Dimerization Interface of the 14-3-3ζ Adapter Protein. In: ChemBioChem: A European Journal of Chemical Biology 19(6), pp. 591–595. doi:10.1002/cbic.201700588

Eisenstein M, Katchalski-Katzir E (2004) On Proteins, Grids, Correlations, and Docking. In: Comptes Rendus Biologies 327(5), pp. 409–420. doi:10.1016/j.crvi.2004.03.006

Elez K, Bonvin AMJJ, Vangone A (2020) Biological vs. Crystallographic Protein Interfaces: An Overview of Computational Approaches for Their Classification. In: Crystals 10(2). doi:10.3390/cryst10020114

Ernst M, Sittel F, Stock G (2015) Contact- And Distance-Based Principal Component Analysis of Protein Dynamics. In: Journal of Chemical Physics 143(24), p. 244114. doi:10.1063/1.4938249

Ernst S, Venkataraman G, Sasisekharan V, Langer R, Cooney CL, Sasisekharan R (1998) Pyranose Ring Flexibility. Mapping of Physical Data for Iduronate in Continuous Conformational Space. In: Journal of the American Chemical Society 120(9), pp. 2099–2107. doi:10.1021/ja972185o

Farshi P, Ohlig S, Pickhinke U, Höing S, Jochmann K, Lawrence R, Dreier R, Dierker T, Grobe K (2011) Dual Roles of the Cardin-Weintraub Motif in Multimeric Sonic Hedgehog. In: Journal of Biological Chemistry 286(26), pp. 23608–23619. doi:10.1074/jbc.M110.206474

Fischer E (1894) Einfluss der Configuration auf die Wirkung der Enzyme. In: Berichte der Deutschen Chemischen Gesellschaft 27(3), pp. 2985–2993. doi:10.1002/cber.18940270364

Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T (2011) Accessing Protein Conformational Ensembles Using Room-Temperature X-Ray Crystallography. In: Proceedings of the National Academy of Sciences of the United States of America 108(39), pp. 16247–16252. doi:10.1073/pnas.1111325108

Free Software Foundation (2007) GNU Lesser General Public License. GNU. Version 3. URL: https://www.gnu.org/licenses/lgpl-3.0.en.html

Freeman AK, Ritt DA, Morrison DK (2013) The Importance of Raf Dimerization in Cell Signaling. In: Small GTPases 4(3), pp. 180–185. doi:10.4161/sgtp.26117

Gabb HA, Jackson RM, Sternberg MJ (1997) Modelling Protein Docking Using Shape Complementarity, Electrostatics and Biochemical Information. In: Journal of Molecular Biology 272(1), pp. 106–120. doi:10.1006/jmbi.1997.1203

Garzon JI, Lopéz-Blanco JR, Pons C, Kovacs J, Abagyan R, Fernandez-Recio J, Chacon P (2009) FRODOCK: A New Approach for Fast Rotational Protein–Protein Docking. In: Bioinformatics (Oxford, England) 25(19), pp. 2544–2551. doi:10.1093/bioinformatics/btp447

Gasteiger J, Marsili M (1978) A New Model for Calculating Atomic Charges in Molecules. In: Tetrahedron Letters 19(34), pp. 3181–3184. doi:10.1016/S0040-4039(01)94977-9

Gasteiger J, Marsili M (1980) Iterative Partial Equalization of Orbital Electronegativity – A Rapid Access to Atomic Charges. In: Tetrahedron 36(22), pp. 3219–3228. doi:10.1016/0040-4020(80)80168-2

Gebala M, Bonilla S, Bisaria N, Herschlag D (2016) Does Cation Size Affect Occupancy and Electrostatic Screening of the Nucleic Acid Ion Atmosphere? In: Journal of the American Chemical Society 138(34), pp. 10925–10934. doi:10.1021/jacs.6b04289

Geidl S, Bouchal T, Raček T, Svobodová Vařeková R, Hejret V, Křenek A, Abagyan R, Koča J (2015) High-Quality and Universal Empirical Atomic Charges for Chemoinformatics Applications. In: Journal of Cheminformatics 7, p. 59. doi:10.1186/s13321-015-0107-1

Gianni S, Dogan J, Jemth P (2014) Distinguishing Induced Fit From Conformational Selection. In: Biophysical Chemistry 189, pp. 33–39. doi:10.1016/j.bpc.2014.03.003

Gigante A, Grad JN, Briels J, Bartel M, Hoffmann D, Ottmann C, Schmuck C (2019) A New Class of Supramolecular Ligands Stabilizes 14-3-3 Protein–Protein Interactions by Up to Two Orders of Magnitude. In: Chemical Communications (Cambridge, England) 55(1), pp. 111–114. doi:10.1039/c8cc07946c

González Á (2010) Measurement of Areas on a Sphere Using Fibonacci and Latitude–Longitude Lattices. In: Mathematical Geosciences 42(1), p. 49. doi:10.1007/s11004-009-9257-x

Goodford PJ (1985) A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. In: Journal of Medicinal Chemistry 28(7), pp. 849–857. doi:10.1021/jm00145a002

Goodsell DS, Olson AJ (1990) Automated Docking of Substrates to Proteins by Simulated Annealing. In: Proteins 8(3), pp. 195–202. doi:10.1002/prot.340080302

Grad JN, Gigante A, Wilms C, Dybowski JN, Ohl L, Ottmann C, Schmuck C, Hoffmann D (2018) Locating Large, Flexible Ligands on Proteins. In: Journal of Chemical Information and Modeling 58(2), pp. 315–327. doi:10.1021/acs.jcim.7b00413

Grad JN, Hoffmann D (2014) Parameterizing Novel Residues for the AMBER99SB Forcefield. In: First CRC1093 Graduate Symposium: Supramolecular Chemistry on Proteins. (Schloss Gnadenthal, Kleve, Germany, Sept. 22, 2014–Sept. 24, 2014)

Grad JN, Hoffmann D (2015) Investigating Multiarmed Peptidic Dendrimers Using MD Simulations. In: CRC1093 International Symposium: Supramolecular Chemistry on Proteins. (University of Duisburg-Essen, Essen, Germany, Sept. 29, 2015–Sept. 30, 2015), p. 11. URL: https://www.uni-due.de/imperia/md/content/crc1093/download/crc1093_international_symposium_2015_-poster_abstract_booklet-sept19.pdf#page=11

Grad JN, Hoffmann D (2016) Computational Investigation of Multivalent Peptidic Dendrimers. In: CRC1093 and CRC765 Graduate Student Symposium: Protein–Ligand Interactions. (GHotel Hannover, Kleefeld, Germany, Aug. 31, 2016–Sept. 2, 2016), p. 31. URL: https://www.uni-due.de/imperia/md/images/crc1093/graduate-school/crc1093_and_crc765_graduate_student_symposium_2016_-_abstract_book_-_aug2016.pdf#page=31

Grad JN, Hoffmann D (2017a) Molecular Dynamics Simulation Steady-State Prediction by Electrostatics – Applications in Supramolecular Chemistry. In: 2nd International Symposium: Supramolecular Chemistry on Proteins. (University of Duisburg-Essen, Essen, Germany, Sept. 20, 2017–Sept. 21, 2017), p. 14. URL: https://www.uni-due.de/imperia/md/content/crc1093/download/crc1093_international_symposium_2017_-poster_abstract_booklet-final.pdf#page=14

Grad JN, Hoffmann D (2017b) Molecular Dynamics Simulation Steady-State Prediction by Electrostatics — Applications in Supramolecular Chemistry. In: 11th Triennial Congress of the World Association of Theoretical and Computational Chemists, Book of Abstracts. (Gasteig Cultural Center, Munich, Germany, Aug. 27, 2017–Sept. 1, 2017), p. 1042. URL: http://www.watoc2017.com/files/WATOC17/Downloads/Book_of_Abstracts_final.pdf#page=1042

Grad JN, Hoffmann D (2017c) Targeting Protein–Protein Interactions via Supramolecular Ligands – Mechanistic Insight by Molecular Modeling. In: Molecular Basis of Life, Abstracts. (Ruhr University Bochum, Bochum, Germany, Sept. 24, 2017–Sept. 27, 2017), p. 67. URL: https://gbm-online.de/tagungskalender-details/molecular-basis-of-life-2017-herbsttagung-der-gbm-25.html?file=files/gbm/tagungen/fall%20meetings/2017_MBoL_Abstracts.pdf#page=67

Grad JN, Weeber R, Holzer M, Holm C (2022) Making Use of Established Codes Instead of Re-Inventing the Wheel: Using Lattice-Based Methods From waLBerla Into ESPResSo. In: SE22 – Software Engineering. (online event, Department of Computer Science, Humboldt-Universität zu Berlin, Germany, Feb. 21, 2022–Feb. 25, 2022). URL: https://pad.gwdg.de/XsfpS3qtSNOgSPcMKIEeCw?view#Making-use-of-established-

codes-instead-of-re-inventing-the-wheel-Using-lattice-based-methods-from-waLBerla-into-ESPResSo

Grad JN, Weeber R, Weik F, Szuttor K, Holm C (2019) ESPResSo: Experiences in Lowering the Entry Barrier for New Developers in a Particle-Based Simulation Package. In: deRSE19 – Conference for Research Software Engineers in Germany. (Albert Einstein Science Park, Potsdam, Germany, June 4, 2019–June 6, 2019). URL: https://derse19.uni-jena.de/derse19/talk/review/SGSKQL8N3FFV9W8TE8PTAFC3RVLWEYX3

Graen T, Hoefling M, Grubmüller H (2014) AMBER-DYES: Characterization of Charge Fluctuations and Force Field Parameterization of Fluorescent Dyes for Molecular Dynamics Simulations. In: Journal of Chemical Theory and Computation 10(12), pp. 5505–5512. doi:10.1021/ct500869p

Gražulis S, Daškevič A, Merkys A, Chateigner D, Lutterotti L, Quirós M, Serebryanaya NR, Moeck P, Downs RT, Le Bail A (2012) Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. In: Nucleic Acids Research 40(Database issue), pp. D420–D427. doi:10.1093/nar/gkr900

Greer J, Bush BL (1978) Macromolecular Shape and Surface Maps by Solvent Exclusion. In: Proceedings of the National Academy of Sciences of the United States of America 75(1), pp. 303–307. doi:10.1073/pnas.75.1.303

Grove LE, Hall DR, Beglov D, Vajda S, Kozakov D (2013) FTFlex: Accounting for Binding Site Flexibility to Improve Fragment-Based Identification of Druggable Hot Spots. In: Bioinformatics (Oxford, England) 29(9), pp. 1218–1219. doi:10.1093/bioinformatics/btt102

Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions. In: Proteins 47(4), pp. 409–443. doi:10.1002/prot.10115

Handing KB, Niedzialkowska E, Shabalin IG, Kuhn ML, Zheng H, Minor W (2018) Characterizing Metal-Binding Sites in Proteins With X-Ray Crystallography. In: Nature Protocols 13(5), pp. 1062–1090. doi:10.1038/nprot.2018.018

Hannay JH, Nye JF (2004) Fibonacci Numerical Integration on a Sphere. In: Journal of Physics A: Mathematical and General 37(48), p. 11591. doi:10.1088/0305-4470/37/48/005

Harrison RW, Kourinov IV, Andrews LC (1994) The Fourier–Green's Function and the Rapid Evaluation of Molecular Potentials. In: Protein Engineering 7(3), pp. 359–369. doi:10.1093/protein/7.3.359

Hartman AM, Hirsch AKH (2017) Molecular Insight Into Specific 14-3-3 Modulators: Inhibitors and Stabilisers of Protein–Protein Interactions of 14-3-3. In: European Journal of Medicinal Chemistry 136, pp. 573–584. doi:10.1016/j.ejmech.2017.04.058

Hartree DR, Hartree W (1935) Self-Consistent Field, With Exchange, for Beryllium. In: Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences 150(869), pp. 9–33. doi:10.1098/rspa.1935.0085

Hatzivassiliou G, Song K, Yen I, Brandhuber BJ, Anderson DJ, Alvarado R, Ludlam MJC, Stokoe D, Gloor SL, Vigers G, Morales T, Aliagas I, Liu B, Sideris S,

Hoeflich KP, Jaiswal BS, Seshagiri S, Koeppen H, Belvin M, Friedman LS, Malek S (2010) RAF Inhibitors Prime Wild-Type RAF to Activate the MAPK Pathway and Enhance Growth. In: Nature 464(7287), pp. 431–435. doi:10.1038/nature08833

Hegde RP, Pavithra GC, Dey D, Almo SC, Ramakumar S, Ramagopal UA (2017) Can the Propensity of Protein Crystallization Be Increased by Using Systematic Screening With Metals? In: Protein Science: A Publication of the Protein Society 26(9), pp. 1704–1713. doi:10.1002/pro.3214

Heidorn SJ, Milagre C, Whittaker S, Nourry A, Niculescu-Duvas I, Dhomen N, Hussain J, Reis-Filho JS, Springer CJ, Pritchard C, Marais R (2010) Kinase-Dead BRAF and Oncogenic RAS Cooperate to Drive Tumor Progression Through CRAF. In: Cell 140(2), pp. 209–221. doi:10.1016/j.cell.2009.12.040

Heifetz A, Katchalski-Katzir E, Eisenstein M (2002) Electrostatics in Protein–Protein Docking. In: Protein Science: A Publication of the Protein Society 11(3), pp. 571–587. doi:10.1110/ps.26002

Hermeking H (2003) The 14-3-3 Cancer Connection. In: Nature Reviews Cancer 3(12), pp. 931–943. doi:10.1038/nrc1230

Hertzman Johansson C, Egyhazi Brage S (2014) BRAF Inhibitors in Cancer Therapy. In: Pharmacology & Therapeutics 142(2), pp. 176–182. doi:10.1016/j.pharmthera.2013.11.011

Heuser P, Baù D, Benkert P, Schomburg D (2005) Refinement of Unbound Protein Docking Studies Using Biological Knowledge. In: Proteins 61(4), pp. 1059–1067. doi:10.1002/prot.20634

Hsieh PH, Thieker DF, Guerrini M, Woods RJ, Liu J (2016) Uncovering the Relationship Between Sulphation Patterns and Conformation of Iduronic Acid in Heparan Sulphate. In: Scientific Reports 6, p. 29602. doi:10.1038/srep29602

Hu L, Benson ML, Smith RD, Lerner MG, Carlson HA (2005) Binding MOAD (Mother of All Databases). In: Proteins 60(3), pp. 333–340. doi:10.1002/prot.20512

Huang T, Karsy M, Zhuge J, Zhong M, Liu D (2013) B-Raf and the Inhibitors: From Bench to Bedside. In: Journal of Hematology & Oncology 6, p. 30. doi:10.1186/1756-8722-6-30

Huerta-Viga A, Amirjalayer S, Domingos SR, Meuzelaar H, Rupenyan A, Woutersen S (2015) The Structure of Salt Bridges Between Arg^+ and Glu^- in Peptides Investigated With 2D-IR Spectroscopy: Evidence for Two Distinct Hydrogen-Bond Geometries. In: Journal of Chemical Physics 142(21), p. 212444. doi:10.1063/1.4921064

Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A Semiempirical Free Energy Force Field With Charge-Based Desolvation. In: Journal of Computational Chemistry 28(6), pp. 1145–1152. doi:10.1002/jcc.20634

Ingham PW, Nakano Y, Seger C (2011) Mechanisms and Functions of Hedgehog Signalling Across the Metazoa. In: Nature Reviews Genetics 12(6), pp. 393–406. doi:10.1038/nrg2984

IUPAC (1983) Symbols for Specifying the Conformation of Polysaccharide Chains. In: European Journal of Biochemistry 131(2), pp. 5–7. doi:10.1111/j.1432-1033.1983.tb07224.x

Jakalian A, Bush BL, Jack DB, Bayly CI (2000) Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. In: Journal of Computational Chemistry 21(2), pp. 132–146. doi:10.1002/(SICI)1096-987X(20000130)21:2<132::AID-JCC5>3.0.CO;2-P

Jakalian A, Jack DB, Bayly CI (2002) Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. In: Journal of Computational Chemistry 23(16), pp. 1623–1641. doi:10.1002/jcc.10128

Jämbeck JPM, Lyubartsev AP (2014) Update to the General Amber Force Field for Small Solutes With an Emphasis on Free Energies of Hydration. In: Journal of Physical Chemistry B 118(14), pp. 3793–3804. doi:10.1021/jp4111234

Janin J (1999) Wet and Dry Interfaces: The Role of Solvent in Protein–Protein and Protein–DNA Recognition. In: Structure (London, England: 1993) 7(12), R277–R279. doi:10.1016/s0969-2126(00)88333-1

Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJE, Vajda S, Vakser I, Wodak SJ, Critical Assessment of PRedicted Interactions (2003) CAPRI: A Critical Assessment of PRedicted Interactions. In: Proteins 52(1), pp. 2–9. doi:10.1002/prot.10381

Jiang F, Kim SH (1991) "Soft Docking": Matching of Molecular Surface Cubes. In: Journal of Molecular Biology 219(1), pp. 79–102. doi:10.1016/0022-2836(91)90859-5

Jiang L, Kuhlman B, Kortemme T, Baker D (2005) A "Solvated Rotamer" Approach to Modeling Water-Mediated Hydrogen Bonds at Protein–Protein Interfaces. In: Proteins 58(4), pp. 893–904. doi:10.1002/prot.20347

Jiang QQ, Bartsch L, Sicking W, Wich PR, Heider D, Hoffmann D, Schmuck C (2013) A New Approach to Inhibit Human β-Tryptase by Protein Surface Binding of Four-Armed Peptide Ligands With Two Different Sets of Arms. In: Organic & Biomolecular Chemistry 11(10), pp. 1631–1639. doi:10.1039/c3ob27302d

Jiang QQ, Sicking W, Ehlers M, Schmuck C (2015) Discovery of Potent Inhibitors of Human β-Tryptase From Pre-Equilibrated Dynamic Combinatorial Libraries. In: Chemical Science 6(3), pp. 1792–1800. doi:10.1039/C4SC02943G

Jiroušková Z, Vařeková RS, Vaněk J, Koča J (2009) Electronegativity Equalization Method: Parameterization and Validation for Organic Molecules Using the Merz-Kollman-Singh Charge Distribution Scheme. In: Journal of Computational Chemistry 30(7), pp. 1174–1178. doi:10.1002/jcc.21142

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. In: The Journal of Chemical Physics 79(2), pp. 926–935. doi:10.1063/1.445869

Jukič M, Konc J, Gobec S, Janežič D (2017) Identification of Conserved Water Sites in Protein Structures for Drug Design. In: Journal of Chemical Information and Modeling 57(12), pp. 3094–3103. doi:10.1021/acs.jcim.7b00443

Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, Chun M, Li P, Gohara DW, Dolinsky T, Konecny R, Koes DR, Nielsen JE, Head-Gordon T, Geng W, Krasny R, Wei GW, Holst MJ, McCammon JA, Baker NA (2018) Improvements to the APBS Biomolecular Solvation

Software Suite. In: Protein Science: A Publication of the Protein Society 27(1), pp. 112–128. doi:10.1002/pro.3280

Kaplan A, Fournier AE (2017) Targeting 14-3-3 Adaptor Protein-Protein Interactions to Stimulate Central Nervous System Repair. In: Neural Regeneration Research 12(7), pp. 1040–1043. doi:10.4103/1673-5374.211176

Kaplan A, Ottmann C, Fournier AE (2017) 14-3-3 Adaptor Protein-Protein Interactions as Therapeutic Targets for CNS Diseases. In: Pharmacological Research 125(Pt B), pp. 114–121. doi:10.1016/j.phrs.2017.09.007

Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA (1992) Molecular Surface Recognition: Determination of Geometric Fit Between Proteins and Their Ligands by Correlation Techniques. In: Proceedings of the National Academy of Sciences of the United States of America 89(6), pp. 2195–2199. doi:10.1073/pnas.89.6.2195

Kendall MG, Moran PAP (1963) Geometrical Probability. Ed. by Kendall MG. Griffin's Statistical Monographs & Courses. Charles Griffin & Company, Limited, London

Khan S, Fung KW, Rodriguez E, Patel R, Gor J, Mulloy B, Perkins SJ (2013) The Solution Structure of Heparan Sulfate Differs From That of Heparin: Implications for Function. In: Journal of Biological Chemistry 288(39), pp. 27737–27751. doi:10.1074/jbc.M113.492223

Khazak V, Astsaturov I, Serebriiskii IG, Golemis EA (2007) Selective Raf Inhibition in Cancer Therapy. In: Expert Opinion on Therapeutic Targets 11(12), pp. 1587–1609. doi:10.1517/14728222.11.12.1587

Kim EK, Choi EJ (2010) Pathological Roles of MAPK Signaling Pathways in Human Diseases. In: Biochimica et Biophysica Acta 1802(4), pp. 396–405. doi:10.1016/j.bbadis.2009.12.009

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by Simulated Annealing. In: Science (New York, N.Y.) 220(4598), pp. 671–680. doi:10.1126/science.220.4598.671

Kirmizialtin S, Silalahi ARJ, Elber R, Fenley MO (2012) The Ionic Atmosphere Around A-RNA: Poisson-Boltzmann and Molecular Dynamics Simulations. In: Biophysical Journal 102(4), pp. 829–838. doi:10.1016/j.bpj.2011.12.055

Kirschner KN, Woods RJ (2001) Solvent Interactions Determine Carbohydrate Conformation. In: Proceedings of the National Academy of Sciences of the United States of America 98(19), pp. 10541–10545. doi:10.1073/pnas.191362798

Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ (2008) GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. In: Journal of Computational Chemistry 29(4), pp. 622–655. doi:10.1002/jcc.20820

Klebe G (2015) Applying Thermodynamic Profiling in Lead Finding and Optimization. In: Nature Reviews Drug Discovery 14(2), pp. 95–110. doi:10.1038/nrd4486

Kleywegt GJ (2000) Validation of Protein Crystal Structures. In: Acta Crystallographica. Section D, Biological Crystallography 56(Pt 3), pp. 249–265. doi:10.1107/S0907444999016364

Kortvelyesi T, Dennis S, Silberstein M, Brown L, Vajda S (2003) Algorithms for Computational Solvent Mapping of Proteins. In: Proteins 51(3), pp. 340–351. doi:10.1002/prot.10287

Koshland Jr. DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. In: Proceedings of the National Academy of Sciences of the United States of America 44(2), pp. 98–104. doi:10.1073/pnas.44.2.98

Kovacs JA, Chacón P, Cong Y, Metwally E, Wriggers W (2003) Fast Rotational Matching of Rigid Bodies by Fast Fourier Transform Acceleration of Five Degrees of Freedom. In: Acta Crystallographica. Section D, Biological Crystallography 59(Pt 8), pp. 1371–1376. doi:10.1107/S0907444903011247

Kovacs JA, Wriggers W (2002) Fast Rotational Matching. In: Acta Crystallographica. Section D, Biological Crystallography 58(Pt 8), pp. 1282–1286. doi:10.1107/S0907444902009794

Kozakov D, Brenke R, Comeau SR, Vajda S (2006) PIPER: An FFT-Based Protein Docking Program With Pairwise Potentials. In: Proteins 65(2), pp. 392–406. doi:10.1002/prot.21117

Kubinyi H (2007) Hydrogen Bonding: The Last Mystery in Drug Design? In: Pharmacokinetic Optimization in Drug Research. John Wiley & Sons, Ltd, pp. 513–524. ISBN:9783906390437. doi:10.1002/9783906390437.ch28

Kumar S, Nussinov R (2002) Relationship Between Ion Pair Geometries and Electrostatic Strengths in Proteins. In: Biophysical Journal 83(3), pp. 1595–1612. doi:10.1016/S0006-3495(02)73929-5

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A Geometric Approach to Macromolecule–Ligand Interactions. In: Journal of Molecular Biology 161(2), pp. 269–288. doi:10.1016/0022-2836(82)90153-x

Lamson DR, Hughes MA, Adcock AF, Smith GR, Williams KP (2021) Development and Validation of a Hedgehog Heparin-Binding Assay for High-Throughput Screening. In: MethodsX 8, p. 101207. doi:10.1016/j.mex.2020.101207

Lang PT, Holton JM, Fraser JS, Alber T (2014) Protein Structural Ensembles Are Revealed by Redefining X-Ray Electron Density Noise. In: Proceedings of the National Academy of Sciences of the United States of America 111(1), pp. 237–242. doi:10.1073/pnas.1302823110

Laskowski RA (2003) Structural Quality Assurance. In: Structural Bioinformatics. John Wiley & Sons, Ltd. Chap. 14, pp. 273–303. ISBN:9780471721208. doi:10.1002/0471721204.ch14

Lattman EE (1972) Optimal Sampling of the Rotation Function. In: Acta Crystallographica Section B 28(4), pp. 1065–1068. doi:10.1107/S0567740872003723

Lavery R, Maddocks JH, Pasi M, Zakrzewska K (2014) Analyzing Ion Distributions Around DNA. In: Nucleic Acids Research 42(12), pp. 8138–8149. doi:10.1093/nar/gku504

Lavigne JJ, Anslyn EV (2001) Sensing a Paradigm Shift in the Field of Molecular Recognition: From Selective to Differential Receptors. In: Angewandte Chemie International Edition in English 40(17), pp. 3118–3130. doi:10.1002/1521-3773(20010903)40:17<3118::AID-ANIE3118>3.0.CO;2-Y

Ledvina PS, Yao N, Choudhary A, Quiocho FA (1996) Negative Electrostatic Surface Potential of Protein Sites Specific for Anionic Ligands. In: Proceedings of the National Academy of Sciences of the United States of America 93(13), pp. 6786–6791. doi:10.1073/pnas.93.13.6786

Li L, Chen R, Weng Z (2003) RDOCK: Refinement of Rigid-Body Protein Docking Predictions. In: Proteins 53(3), pp. 693–707. doi:10.1002/prot.10460

Li Y, Zhang H, Litingtung Y, Chiang C (2006) Cholesterol Modification Restricts the Spread of Shh Gradient in the Limb Bud. In: Proceedings of the National Academy of Sciences of the United States of America 103(17), pp. 6548–6553. doi:10.1073/pnas.0600124103

Lindemann SR, Yershova A, LaValle SM (2005) Incremental Grid Sampling Strategies in Robotics. In: Algorithmic Foundations of Robotics VI. Ed. by Erdmann M, Overmars M, Hsu D, van der Stappen F. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 313–328. ISBN:978-3-540-31506-3. doi:10.1007/10991541_22

Liu D, Bienkowska J, Petosa C, Collier RJ, Fu H, Liddington R (1995) Crystal Structure of the Zeta Isoform of the 14-3-3 Protein. In: Nature 376(6536), pp. 191–194. doi:10.1038/376191a0

Luo ZJ, Zhang XF, Rapp U, Avruch J (1995) Identification of the 14.3.3 ζ Domains Important for Self-Association and Raf Binding. In: Journal of Biological Chemistry 270(40), pp. 23681–23687. doi:10.1074/jbc.270.40.23681

Mackay AL, Finney JL, Gotoh K (1977) The Closest Packing of Equal Spheres on a Spherical Surface. In: Acta Crystallographica Section A 33(1), pp. 98–100. doi:10.1107/S0567739477000205

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FT, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. In: Journal of Physical Chemistry B 102(18), pp. 3586–3616. doi:10.1021/jp973084f

Mackintosh C (2004) Dynamic Interactions Between 14-3-3 Proteins and Phosphoproteins Regulate Diverse Cellular Processes. In: Biochemical Journal 381(Pt 2), pp. 329–342. doi:10.1042/BJ20031332

Mahoney MW, Jorgensen WL (2000) A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. In: The Journal of Chemical Physics 112(20), pp. 8910–8922. doi:10.1063/1.481505

Mandell JG, Roberts VA, Pique ME, Kotlovyi V, Mitchell JC, Nelson E, Tsigelny I, Ten Eyck LF (2001) Protein Docking Using Continuum Electrostatics and Geometric Fit. In: Protein Engineering 14(2), pp. 105–113. doi:10.1093/protein/14.2.105

Manikowski D, Jakobs P, Jboor H, Grobe K (2019) Soluble Heparin and Heparan Sulfate Glycosaminoglycans Interfere With Sonic Hedgehog Solubilization and Receptor Binding. In: Molecules (Basel, Switzerland) 24(8). doi:10.3390/molecules24081607

Matta A, Siu KWM, Ralhan R (2012) 14-3-3 Zeta as Novel Molecular Target for Cancer Therapy. In: Expert Opinion on Therapeutic Targets 16(5), pp. 515–523. doi:10.1517/14728222.2012.668185

Maximova T, Moffatt R, Ma B, Nussinov R, Shehu A (2016) Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. In: PLoS Computational Biology 12(4), e1004619. doi:10.1371/journal.pcbi.1004619

McCubrey JA, Steelman LS, Abrams SL, Lee JT, Chang F, Bertrand FE, Navolanic PM, Terrian DM, Franklin RA, D'Assoro AB, Salisbury JL, Mazzarino MC, Stivala F, Libra M (2006) Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT Pathways in Malignant Transformation and Drug Resistance. In: Advances in Enzyme Regulation 46, pp. 249–279. doi:10.1016/j.advenzreg.2006.01.004

McCubrey JA, Steelman LS, Chappell WH, Abrams SL, Wong EWT, Chang F, Lehmann B, Terrian DM, Milella M, Tafuri A, Stivala F, Libra M, Basecke J, Evangelisti C, Martelli AM, Franklin RA (2007) Roles of the Raf/MEK/ERK Pathway in Cell Growth, Malignant Transformation and Drug Resistance. In: Biochimica et Biophysica Acta 1773(8), pp. 1263–1284. doi:10.1016/j.bbamcr.2006.10.001

McPherson A (2017) Protein Crystallization. In: Protein Crystallography: Methods and Protocols. Ed. by Wlodawer A, Dauter Z, Jaskolski M. Springer New York, New York, NY, pp. 17–50. ISBN:978-1-4939-7000-1. doi:10.1007/978-1-4939-7000-1_2

McPhillips F, Mullen P, MacLeod KG, Sewell JM, Monia BP, Cameron DA, Smyth JF, Langdon SP (2006) Raf-1 Is the Predominant Raf Isoform That Mediates Growth Factor-Stimulated Growth in Ovarian Cancer Cells. In: Carcinogenesis 27(4), pp. 729–739. doi:10.1093/carcin/bgi289

Medeiros DdJ, Cortopassi WA, França TCC, Pimentel AS (2013) ITP Adjuster 1.0: A New Utility Program to Adjust Charges in the Topology Files Generated by the PRODRG Server. In: Journal of Chemistry 2013, pp. 1–6. doi:10.1155/2013/803151

Meiners A, Bäcker S, Hadrović I, Heid C, Beuck C, Ruiz-Blanco YB, Mieres-Perez J, Pörschke M, Grad JN, Vallet C, Hoffmann D, Bayer P, Sánchez-García E, Schrader T, Knauer SK (2021) Specific Inhibition of the Survivin–CRM1 Interaction by Peptide-Modified Molecular Tweezers. In: Nature Communications 12(1), p. 1505. doi:10.1038/s41467-021-21753-9

Meyer M, Wilson P, Schomburg D (1996) Hydrogen Bonding and Molecular Surface Shape Complementarity as a Basis for Protein Docking. In: Journal of Molecular Biology 264(1), pp. 199–210. doi:10.1006/jmbi.1996.0634

Miles RE (1965) On Random Rotations in R^3. In: Biometrika 52(3-4), pp. 636–639. doi:10.1093/biomet/52.3-4.636

Millman KJ, Aivazis M (2011) Python for Scientists and Engineers. In: Computing in Science & Engineering 13(2), pp. 9–12. doi:10.1109/MCSE.2011.36

Milroy LG, Bartel M, Henen MA, Leysen S, Adriaans JMC, Brunsveld L, Landrieu I, Ottmann C (2015) Stabilizer-Guided Inhibition of Protein–Protein Interactions. In: Angewandte Chemie International Edition in English 54(52), pp. 15720–15724. doi:10.1002/anie.201507976

Mintseris J, Pierce B, Wiehe K, Anderson R, Chen R, Weng Z (2007) Integrating Statistical Pair Potentials Into Protein Complex Prediction. In: Proteins 69(3), pp. 511–520. doi:10.1002/prot.21502

Miyazawa S, Jernigan RL (1985) Estimation of Effective Interresidue Contact Energies From Protein Crystal Structures: Quasi-Chemical Approximation. In: Macromolecules 18(3), pp. 534–552. doi:10.1021/ma00145a039

Molzan M, Kasper S, Röglin L, Skwarczynska M, Sassa T, Inoue T, Breitenbuecher F, Ohkanda J, Kato N, Schuler M, Ottmann C (2013) Stabilization of Physical RAF/14-3-3 Interaction by Cotylenin A as Treatment Strategy for RAS Mutant Cancers. In: ACS Chemical Biology 8(9), pp. 1869–1875. doi:10.1021/cb4003464

Molzan M, Ottmann C (2012) Synergistic Binding of the Phosphorylated S233- And S259-Binding Sites of C-RAF to One 14-3-3ζ Dimer. In: Journal of Molecular Biology 423(4), pp. 486–495. doi:10.1016/j.jmb.2012.08.009

Momany FA (1978) Determination of Partial Atomic Charges From *Ab Initio* Molecular Electrostatic Potentials. Application to Formamide, Methanol, and Formic Acid. In: Journal of Physical Chemistry 82(5), pp. 592–601. doi:10.1021/j100494a019

Montagut C, Settleman J (2009) Targeting the RAF-MEK-ERK Pathway in Cancer Therapy. In: Cancer Letters 283(2), pp. 125–134. doi:10.1016/j.canlet.2009.01.022

Moont G, Gabb HA, Sternberg MJ (1999) Use of Pair Potentials Across Protein Interfaces in Screening Predicted Docked Complexes. In: Proteins 35(3), pp. 364–373. doi:10.1002/(SICI)1097-0134(19990515)35:3<364::AID-PROT11>3.0.CO;2-4

Moreira IS, Fernandes PA, Ramos MJ (2010) Protein–Protein Docking Dealing With the Unknown. In: Journal of Computational Chemistry 31(2), pp. 317–342. doi:10.1002/jcc.21276

Mori M, Vignaroli G, Botta M (2013) Small Molecules Modulation of 14-3-3 Protein–Protein Interactions. In: Drug Discovery Today: Technologies 10(4), e541–e547. doi:10.1016/j.ddtec.2012.10.001

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated Docking With Selective Receptor Flexibility. In: Journal of Computational Chemistry 30(16), pp. 2785–2791. doi:10.1002/jcc.21256

Morrison JL, Breitling R, Higham DJ, Gilbert DR (2006) A Lock-and-Key Model for Protein–Protein Interactions. In: Bioinformatics (Oxford, England) 22(16), pp. 2012–2019. doi:10.1093/bioinformatics/btl338

Mortier WJ, Van Genechten K, Gasteiger J (1985) Electronegativity Equalization: Application and Parametrization. In: Journal of the American Chemical Society 107(4), pp. 829–835. doi:10.1021/ja00290a017

Mortier WJ, Ghosh SK, Shankar S (1986) Electronegativity-Equalization Method for the Calculation of Atomic Charges in Molecules. In: Journal of the American Chemical Society 108(15), pp. 4315–4320. doi:10.1021/ja00275a013

Mottarella SE, Beglov D, Beglova N, Nugent MA, Kozakov D, Vajda S (2014) Docking Server for the Identification of Heparin Binding Sites on Proteins. In: Journal of Chemical Information and Modeling 54(7), pp. 2068–2078. doi:10.1021/ci500115j

Mulloy B, Forster MJ (2000) Conformation and Dynamics of Heparin and Heparan Sulfate. In: Glycobiology 10(11), pp. 1147–1156. doi:10.1093/glycob/10.11.1147

Muñoz-García JC, Corzana F, de Paz JL, Angulo J, Nieto PM (2013) Conformations of the Iduronate Ring in Short Heparin Fragments Described by Time-Averaged Distance Restrained Molecular Dynamics. In: Glycobiology 23(11), pp. 1220–1229. doi:10.1093/glycob/cwt058

Nagy G, Oostenbrink C, Hritz J (2017) Exploring the Binding Pathways of the 14-3-3ζ Protein: Structural and Free-Energy Profiles Revealed by Hamiltonian Replica Exchange Molecular Dynamics With Distancefield Distance Restraints. In: PLoS One 12(7), e0180633. doi:10.1371/journal.pone.0180633

Nannenga BL (2020) MicroED Methodology and Development. In: Structural Dynamics 7(1), p. 014304. doi:10.1063/1.5128226

Neal CL, Yu D (2010) 14-3-3ζ as a Prognostic Marker and Therapeutic Target for Cancer. In: Expert Opinion on Therapeutic Targets 14(12), pp. 1343–1354. doi:10.1517/14728222.2010.531011

Ngan CH, Bohnuud T, Mottarella SE, Beglov D, Villar EA, Hall DR, Kozakov D, Vajda S (2012) FTMAP: Extended Protein Mapping With User-Selected Probe Molecules. In: Nucleic Acids Research 40(Web Server issue), W271–W275. doi:10.1093/nar/gks441

O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: An Open Chemical Toolbox. In: Journal of Cheminformatics 3, p. 33. doi:10.1186/1758-2946-3-33

O'Dell WB, Bodenheimer AM, Meilleur F (2016) Neutron Protein Crystallography: A Complementary Tool for Locating Hydrogens in Proteins. In: Archives of Biochemistry and Biophysics 602, pp. 48–60. doi:10.1016/j.abb.2015.11.033

Oliphant TE (2006) Guide to NumPy. 1st ed. Trelgol Publishing, USA

Oliphant TE (2007) Python for Scientific Computing. In: Computing in Science & Engineering 9(3), pp. 10–20. doi:10.1109/MCSE.2007.58

Papoian GA, Ulander J, Wolynes PG (2003) Role of Water Mediated Interactions in Protein–Protein Recognition Landscapes. In: Journal of the American Chemical Society 125(30), pp. 9170–9178. doi:10.1021/ja034729u

Parker MW (2003) Protein Structure From X-Ray Diffraction. In: Journal of Biological Physics 29(4), pp. 341–362. doi:10.1023/A:1027310719146

Pathi S, Pagan-Westphal S, Baker DP, Garber EA, Rayhorn P, Bumcrot D, Tabin CJ, Blake Pepinsky R, Williams KP (2001) Comparative Biological Responses to Human

Sonic, Indian, and Desert Hedgehog. In: Mechanisms of Development 106(1-2), pp. 107–117. doi:10.1016/s0925-4773(01)00427-0

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É (2011) Scikit-Learn: Machine Learning in Python. In: Journal of Machine Learning Research 12, pp. 2825–2830. URL: https://jmlr.org/papers/v12/pedregosa11a.html

Pellegrini S, Grad JN, Bousquet T, Pélinski L (2011) A Novel Multicomponent Reaction: Easy Access to Ferrocenyl (Alkylimino)-1,4-Dihydroquinolines. In: Tetrahedron Letters 52(15), pp. 1742–1744. doi:10.1016/j.tetlet.2011.01.144

Pepinsky RB, Rayhorn P, Day ES, Dergay A, Williams KP, Galdes A, Taylor FR, Boriack-Sjodin PA, Garber EA (2000) Mapping Sonic Hedgehog-Receptor Interactions by Steric Interference. In: Journal of Biological Chemistry 275(15), pp. 10995–11001. doi:10.1074/jbc.275.15.10995

Perez JJ, Tomas MS, Rubio-Martinez J (2016) Assessment of the Sampling Performance of Multiple-Copy Dynamics versus a Unique Trajectory. In: Journal of Chemical Information and Modeling 56(10), pp. 1950–1962. doi:10.1021/acs.jcim.6b00347

Raček T, Pazúriková J, Svobodová Vařeková R, Geidl S, Křenek A, Falginella FL, Horský V, Hejret V, Koča J (2016) NEEMP: Software for Validation, Accurate Calculation and Fast Parameterization of EEM Charges. In: Journal of Cheminformatics 8, p. 57. doi:10.1186/s13321-016-0171-1

Rafieiolhosseini N, Killa M, Neumann T, Tötsch N, Grad JN, Höing A, Dirksmeyer T, Niemeyer J, Ottmann C, Knauer SK, Giese M, Voskuhl J, Hoffmann D (2022) Computational Model Predicts Protein Binding Sites of a Luminescent Ligand Equipped With Guanidiniocarbonyl-Pyrrole Groups. In: Beilstein Journal of Organic Chemistry 18, pp. 1322–1331. doi:10.3762/bjoc.18.137

Rajakulendran T, Adam DN (2014) Bench to Bedside: Mechanistic Principles of Targeting the RAF Kinase in Melanoma. In: International Journal of Dermatology 53(12), pp. 1428–1433. doi:10.1111/ijd.12724

Rajakulendran T, Sahmi M, Lefrançois M, Sicheri F, Therrien M (2009) A Dimerization-Dependent Mechanism Drives RAF Catalytic Activation. In: Nature 461(7263), pp. 542–545. doi:10.1038/nature08314

Reynolds CA, Essex JW, Richards WG (1992) Atomic Charges for Variable Molecular Conformations. In: Journal of the American Chemical Society 114(23), pp. 9075–9079. doi:10.1021/ja00049a045

Riddle RD, Johnson RL, Laufer E, Tabin C (1993) *Sonic hedgehog* Mediates the Polarizing Activity of the ZPA. In: Cell 75(7), pp. 1401–1416. doi:10.1016/0092-8674(93)90626-2

Ritchie DW, Kemp GJ (2000) Protein Docking Using Spherical Polar Fourier Correlations. In: Proteins 39(2), pp. 178–194. doi:10.1002/(SICI)1097-0134(20000501)39:2<178::AID-PROT8>3.0.CO;2-6

Ritchie DW, Grudinin S (2016) Spherical Polar Fourier Assembly of Protein Complexes With Arbitrary Point Group Symmetry. In: Journal of Applied Crystallography 49(1), pp. 158–167. doi:10.1107/S1600576715022931

Ritchie DW, Kozakov D, Vajda S (2008) Accelerating and Focusing Protein–Protein Docking Correlations Using Multi-Dimensional Rotational FFT Generating Functions. In: Bioinformatics (Oxford, England) 24(17), pp. 1865–1873. doi:10.1093/bioinformatics/btn334

Ritchie DW, Venkatraman V (2010) Ultra-Fast FFT Protein Docking on Graphics Processors. In: Bioinformatics (Oxford, England) 26(19), pp. 2398–2405. doi:10.1093/bioinformatics/btq444

Robbins DJ, Fei DL, Riobo NA (2012) The Hedgehog Signal Transduction Network. In: Science Signaling 5(246), re6. doi:10.1126/scisignal.2002906

Rosenquist M, Sehnke P, Ferl RJ, Sommarin M, Larsson C (2000) Evolution of the 14-3-3 Protein Family: Does the Large Number of Isoforms in Multicellular Organisms Reflect Functional Specificity? In: Journal of Molecular Evolution 51(5), pp. 446–458. doi:10.1007/s002390010107

Rubin JB, Choi Y, Segal RA (2002) Cerebellar Proteoglycans Regulate Sonic Hedgehog Responses During Development. In: Development (Cambridge, England) 129(9), pp. 2223–2232. doi:10.1242/dev.129.9.2223

Rühle V, Kusumaatmaja H, Chakrabarti D, Wales DJ (2013) Exploring Energy Landscapes: Metrics, Pathways, and Normal-Mode Analysis for Rigid-Body Molecules. In: Journal of Chemical Theory and Computation 9(9), pp. 4026–4034. doi:10.1021/ct400403y

Rummler H (2002) On the Distribution of Rotation Angles How Great Is the Mean Rotation Angle of a Random Rotation? In: The Mathematical Intelligencer 24(4), pp. 6–11. doi:10.1007/BF03025318

Rushworth LK, Hindley AD, O'Neill E, Kolch W (2006) Regulation and Role of Raf-1/B-Raf Heterodimerization. In: Molecular and Cellular Biology 26(6), pp. 2262–2272. doi:10.1128/MCB.26.6.2262-2272.2006

Russo Krauss I, Merlino A, Vergara A, Sica F (2013) An Overview of Biological Macromolecule Crystallization. In: International Journal of Molecular Sciences 14(6), pp. 11643–11691. doi:10.3390/ijms140611643

Sabarinathan R, Aishwarya K, Sarani R, Vaishnavi MK, Sekar K (2011) Water-Mediated Ionic Interactions in Protein Structures. In: Journal of Biosciences (New Delhi, India) 36(2), pp. 253–263. doi:10.1007/s12038-011-9067-4

Sadourny R (1972) Conservative Finite-Difference Approximations of the Primitive Equations on Quasi-Uniform Spherical Grids. In: Monthly Weather Review 100(2), pp. 136–144. doi:10.1175/1520-0493(1972)100<0136:CFAOTP>2.3.CO;2

Saff EB, Kuijlaars ABJ (1997) Distributing Many Points on a Sphere. In: The Mathematical Intelligencer 19(1), pp. 5–11. doi:10.1007/BF03024331

Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC (2013) Routine Microsecond Molecular Dynamics Simulations With AMBER on GPUs. 2. Explicit Solvent

Particle Mesh Ewald. In: Journal of Chemical Theory and Computation 9(9), pp. 3878–3888. doi:10.1021/ct400314y

Santos-Martins D, Forli S, Ramos MJ, Olson AJ (2014) AutoDock4(Zn): An Improved AutoDock Force Field for Small-Molecule Docking to Zinc Metalloproteins. In: Journal of Chemical Information and Modeling 54(8), pp. 2371–2379. doi:10.1021/ci500209e

Sattelle BM, Hansen SU, Gardiner J, Almond A (2010) Free Energy Landscapes of Iduronic Acid and Related Monosaccharides. In: Journal of the American Chemical Society 132(38), pp. 13132–13134. doi:10.1021/ja1054143

Sauter C, Ng JD, Lorber B, Keith G, Brion P, Hosseini MW, Lehn JM, Giegé R (1999) Additives for the Crystallization of Proteins and Nucleic Acids. In: Journal of Crystal Growth 196(2), pp. 365–376. doi:10.1016/S0022-0248(98)00852-5

Schneider HJ (2009) Binding Mechanisms in Supramolecular Complexes. In: Angewandte Chemie International Edition in English 48(22), pp. 3924–3977. doi:10.1002/anie.200802947

Schneider HJ (2015) Limitations and Extensions of the Lock-and-Key Principle: Differences Between Gas State, Solution and Solid State Structures. In: International Journal of Molecular Sciences 16(4), pp. 6694–6717. doi:10.3390/ijms16046694

Schüttelkopf AW, van Aalten DMF (2004) *PRODRG*: A Tool for High-Throughput Crystallography of Protein–Ligand Complexes. In: Acta Crystallographica. Section D, Biological Crystallography 60(Pt 8), pp. 1355–1363. doi:10.1107/S0907444904011679

Shepherd C, Puzanov I, Sosman JA (2010) B-RAF Inhibitors: An Evolving Role in the Therapy of Malignant Melanoma. In: Current Oncology Reports 12(3), pp. 146–152. doi:10.1007/s11912-010-0095-2

Silhan J, Obsilova V, Vecer J, Herman P, Sulc M, Teisinger J, Obsil T (2004) 14-3-3 Protein *C*-Terminal Stretch Occupies Ligand Binding Groove and Is Displaced by Phosphopeptide Binding. In: Journal of Biological Chemistry 279(47), pp. 49113–49119. doi:10.1074/jbc.M408671200

Singh A, Tessier MB, Pederson K, Wang X, Venot AP, Boons GJ, Prestegard JH, Woods RJ (2016) Extension and Validation of the GLYCAM Force Field Parameters for Modeling Glycosaminoglycans. In: Canadian Journal of Chemistry 94(11), pp. 927–935. doi:10.1139/cjc-2015-0606

Singh UC, Kollman PA (1984) An Approach to Computing Electrostatic Charges for Molecules. In: Journal of Computational Chemistry 5(2), pp. 129–145. doi:10.1002/jcc.540050204

Skjevik ÅA, Madej BD, Walker RC, Teigen K (2012) LIPID11: A Modular Framework for Lipid Simulations Using Amber. In: Journal of Physical Chemistry B 116(36), pp. 11124–11136. doi:10.1021/jp3059992

Smith GR, Sternberg MJE (2003) Evaluation of the 3D-Dock Protein Docking Suite in Rounds 1 and 2 of the CAPRI Blind Trial. In: Proteins 52(1), pp. 74–79. doi:10.1002/prot.10396

Smolin D, Tötsch N, Grad JN, Linders J, Kaschani F, Kaiser M, Kirsch M, Hoffmann D, Schrader T (2020) Accelerated Trypsin Autolysis by Affinity Polymer Templates. In: RSC Advances 10(48), pp. 28711–28719. doi:10.1039/d0ra05827k

Smyth MS, Martin JH (2000) *x* Ray Crystallography. In: Molecular Pathology 53(1), pp. 8–14. doi:10.1136/mp.53.1.8

Sternberg MJE, Aloy P, Gabb HA, Jackson RM, Moont G, Querol E, Avilés FX (1998) A Computational System for Modeling Flexible Protein-Protein and Protein-DNA Docking. In: Glasgow JI, Littlejohn TG, Major F, Lathrop RH, Sankoff D, Sensen C (Ed.) Proceedings of the 6th International Conference on Intelligent Systems for Molecular Biology. (Montreal, Quebec, Canada, June 28, 1998–July 1, 1998). AAAI, Montreal, Quebec, Canada, pp. 183–192. ISBN:1-57735-053-7. URL: https://www.aaai.org/Library/ISMB/1998/ismb98-022.php

Stockham JTG (1966) High-Speed Convolution and Correlation. In: Proceedings of the April 26-28, 1966, Spring Joint Computer Conference. AFIPS '66 (Spring). ACM, Boston, Massachusetts, pp. 229–233. doi:10.1145/1464182.1464209

Stouch TR, Williams DE (1992) Conformational Dependence of Electrostatic Potential Derived Charges of a Lipid Headgroup: Glycerylphosphorylcholine. In: Journal of Computational Chemistry 13(5), pp. 622–632. doi:10.1002/jcc.540130512

Sukharev AG (1971) Optimal Strategies of the Search for an Extremum. In: USSR Computational Mathematics and Mathematical Physics 11(4), pp. 119–137. doi:10.1016/0041-5553(71)90008-5

Sukhwani B, Herbordt MC (2010) FPGA Acceleration of Rigid-Molecule Docking Codes. In: IET Computers & Digital Techniques 4(3), pp. 184–195. doi:10.1049/iet-cdt.2009.0013

Swinbank R, James Purser R (2006) Fibonacci Grids: A Novel Approach to Global Modelling. In: Quarterly Journal of the Royal Meteorological Society 132(619), pp. 1769–1793. doi:10.1256/qj.05.227

Szuttor K, Weik F, Grad JN, Holm C (2021) Modeling the Current Modulation of Bundled DNA Structures in Nanopores. In: The Journal of Chemical Physics 154(5), p. 054901. doi:10.1063/5.0038530

Taylor GL (2010) Introduction to Phasing. In: Acta Crystallographica. Section D, Biological Crystallography 66(Pt 4), pp. 325–338. doi:10.1107/S0907444910006694

Tollenaere JP (1996) The Role of Structure-Based Ligand Design and Molecular Modelling in Drug Discovery. In: Pharmacy World & Science 18(2), pp. 56–62. doi:10.1007/BF00579706

Tovchigrechko A, Vakser IA (2005) Development and Testing of an Automated Approach to Protein Docking. In: Proteins 60(2), pp. 296–301. doi:10.1002/prot.20573

Tovchigrechko A, Vakser IA (2006) GRAMM-X Public Web Server for Protein–Protein Docking. In: Nucleic Acids Research 34(Web Server issue), W310–W314. doi:10.1093/nar/gkl206

Tzivion G, Luo Z, Avruch J (1998) A Dimeric 14-3-3 Protein Is an Essential Cofactor for Raf Kinase Activity. In: Nature 394(6688), pp. 88–92. doi:10.1038/27938

Vajda S, Camacho CJ (2004) Protein–Protein Docking: Is the Glass Half-Full or Half-Empty? In: Trends in Biotechnology 22(3), pp. 110–116. doi:10.1016/j.tibtech.2004.01.006

Vakser IA (1995) Protein Docking for Low-Resolution Structures. In: Protein Engineering 8(4), pp. 371–377. doi:10.1093/protein/8.4.371

Vakser IA (1997) Evaluation of GRAMM Low-Resolution Docking Methodology on the Hemagglutinin-Antibody Complex. In: Proteins Suppl 1, pp. 226–230. doi:10.1002/(SICI)1097-0134(1997)1+<226::AID-PROT31>3.0.CO;2-O

Vakser IA, Aflalo C (1994) Hydrophobic Docking: A Proposed Enhancement to Molecular Recognition Techniques. In: Proteins 20(4), pp. 320–329. doi:10.1002/prot.340200405

Vakser IA, Nikiforovich GV (1995) Protein Docking in the Absence of Detailed Molecular Structures. In: Methods in Protein Structure Analysis. Ed. by Atassi MZ, Appella E. Springer US, Boston, MA, pp. 505–514. ISBN:978-1-4899-1031-8. doi:10.1007/978-1-4899-1031-8_46

Vakser IA (1996) Low-Resolution Docking: Prediction of Complexes for Underdetermined Structures. In: Biopolymers 39(3), pp. 455–464. doi:10.1002/(SICI)1097-0282(199609)39:3<455::AID-BIP16>3.0.CO;2-A

Varga A, Ehrenreiter K, Aschenbrenner B, Kocieniewski P, Kochanczyk M, Lipniacki T, Baccarini M (2017) RAF1/BRAF Dimerization Integrates the Signal From RAS to ERK and ROKα. In: Science Signaling 10(469). doi:10.1126/scisignal.aai8482

Verlinde CL, Hol WG (1994) Structure-Based Drug Design: Progress, Results and Challenges. In: Structure (London, England: 1993) 2(7), pp. 577–587. doi:10.1016/s0969-2126(00)00060-5

Wang J, Cieplak P, Kollman PA (2000) How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? In: Journal of Computational Chemistry 21(12), pp. 1049–1074. doi:10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F

Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and Testing of a General Amber Force Field. In: Journal of Computational Chemistry 25(9), pp. 1157–1174. doi:10.1002/jcc.20035

Wang Q, Shi X, Zhu X, Ehlers M, Wu J, Schmuck C (2014) A Fluorescent Light-Up Probe as an Inhibitor of Intracellular β-Tryptase. In: Chemical Communications (Cambridge, England) 50(46), pp. 6120–6122. doi:10.1039/c4cc02208d

Wang R, Fang X, Lu Y, Yang CY, Wang S (2005) The PDBbind Database: Methodologies and Updates. In: Journal of Medicinal Chemistry 48(12), pp. 4111–4119. doi:10.1021/jm048957q

Wang W, Shakes DC (1996) Molecular Evolution of the 14-3-3 Protein Family. In: Journal of Molecular Evolution 43(4), pp. 384–398. doi:10.1007/BF02339012

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. In: Journal of the American Chemical Society 106(3), pp. 765–784. doi:10.1021/ja00315a051

Whalen DM, Malinauskas T, Gilbert RJC, Siebold C (2013) Structural Insights Into Proteoglycan-Shaped Hedgehog Signaling. In: Proceedings of the National Academy of

Sciences of the United States of America 110(41), pp. 16420–16425. doi:10.1073/pnas.1310097110

Wilker E, Yaffe MB (2004) 14-3-3 Proteins–A Focus on Cancer and Human Disease. In: Journal of Molecular and Cellular Cardiology 37(3), pp. 633–642. doi:10.1016/j.yjmcc.2004.04.015

Williams DE (1973) Optimally Spaced Rotational Grid Points. In: Acta Crystallographica Section A 29(4), pp. 408–414. doi:10.1107/S0567739473001014

Williams DM, Ecroyd H, Goodwin KL, Dai H, Fu H, Woodcock JM, Zhang L, Carver JA (2011) NMR Spectroscopy of 14-3-3ζ Reveals a Flexible C-Terminal Extension: Differentiation of the Chaperone and Phosphoserine-Binding Activities of 14-3-3ζ. In: Biochemical Journal 437(3), pp. 493–503. doi:10.1042/BJ20102178

Williams DE (1990) Alanyl Dipeptide Potential-Derived Net Atomic Charges and Bond Dipoles, and Their Variation With Molecular Conformation. In: Biopolymers 29(10-11), pp. 1367–1386. doi:10.1002/bip.360291005

Williams MA, Goodfellow JM, Thornton JM (1994) Buried Waters and Internal Cavities in Monomeric Proteins. In: Protein Science: A Publication of the Protein Society 3(8), pp. 1224–1235. doi:10.1002/pro.5560030808

Williamson DL (1970) Integration of the Primitive Barotropic Model Over a Spherical Geodesic Grid. In: Monthly Weather Review 98(7), pp. 512–520. doi:10.1175/1520-0493(1970)098<0512:IOTPBM>2.3.CO;2

Williamson DL (1992) Review of Numerical Approaches for Modeling Global Transport. In: Air Pollution Modeling and Its Application IX. Ed. by van Dop H, Kallos G. Springer US, Boston, MA, pp. 377–394. ISBN:978-1-4615-3052-7. doi:10.1007/978-1-4615-3052-7_38

Wilms C (2013) Methods for the Prediction of Complex Biomolecular Structures. PhD thesis. Bioinformatics group, Faculty of Biology, University of Duisburg–Essen. URL: https://duepublico.uni-duisburg-essen.de/servlets/DocumentServlet?id=33166

Wlodawer A, Minor W, Dauter Z, Jaskolski M (2008) Protein Crystallography for Non-Crystallographers, or How to Get the Best (But Not More) From Published Macromolecular Structures. In: FEBS Journal 275(1), pp. 1–21. doi:10.1111/j.1742-4658.2007.06178.x

Woodcock JM, Coolen C, Goodwin KL, Baek DJ, Bittman R, Samuel MS, Pitson SM, Lopez AF (2015) Destabilisation of Dimeric 14-3-3 Proteins as a Novel Approach to Anti-Cancer Therapeutics. In: Oncotarget 6(16), pp. 14522–14536. doi:10.18632/oncotarget.3995

Woods RJ, Chappelle R (2000) Restrained Electrostatic Potential Atomic Partial Charges for Condensed-Phase Simulations of Carbohydrates. In: Theochem 527(1-3), pp. 149–156. doi:10.1016/S0166-1280(00)00487-5

Woods RJ, Dwek RA, Edge CJ, Fraser-Reid B (1995) Molecular Mechanical and Molecular Dynamic Simulations of Glycoproteins and Oligosaccharides. 1. GLYCAM_93 Parameter Development. In: Journal of Physical Chemistry 99(11), pp. 3832–3846. doi:10.1021/j100011a061

Wüthrich K (1995) NMR - This Other Method for Protein and Nucleic Acid Structure Determination. In: Acta Crystallographica. Section D, Biological Crystallography 51(Pt 3), pp. 249–270. doi:10.1107/S0907444994010188

Yaffe MB, Rittinger K, Volinia S, Caron PR, Aitken A, Leffers H, Gamblin SJ, Smerdon SJ, Cantley LC (1997) The Structural Basis for 14-3-3:phosphopeptide Binding Specificity. In: Cell 91(7), pp. 961–971. doi:10.1016/s0092-8674(00)80487-0

Yan S, Peck JM, Ilgu M, Nilsen-Hamilton M, Lamm MH (2020) Sampling Performance of Multiple Independent Molecular Dynamics Simulations of an RNA Aptamer. In: ACS Omega 5(32), pp. 20187–20201. doi:10.1021/acsomega.0c01867

Yang X, Cao W, Zhang L, Zhang W, Zhang X, Lin H (2012) Targeting 14-3-3zeta in Cancer Therapy. In: Cancer Gene Therapy 19(3), pp. 153–159. doi:10.1038/cgt.2011.85

Yershova A, LaValle SM (2004) Deterministic Sampling Methods for Spheres and $SO(3)$. In: IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004. Vol. 4, pp. 3974–3980. doi:10.1109/ROBOT.2004.1308891

Yip-Schneider MT, Miao W, Lin A, Barnard DS, Tzivion G, Marshall MS (2000) Regulation of the Raf-1 Kinase Domain by Phosphorylation and 14-3-3 Association. In: Biochemical Journal 351(Pt 1), pp. 151–159. doi:10.1042/bj3510151

Zebisch A, Troppmair J (2006) Back to the Roots: The Remarkable RAF Oncogene Story. In: Cellular and Molecular Life Sciences: CMLS 63(11), pp. 1314–1330. doi:10.1007/s00018-006-6005-y

Zhang C, Vasmatzis G, Cornette JL, DeLisi C (1997) Determination of Atomic Desolvation Energies From the Structures of Crystallized Proteins. In: Journal of Molecular Biology 267(3), pp. 707–726. doi:10.1006/jmbi.1996.0859

Zhang F, McLellan JS, Ayala AM, Leahy DJ, Linhardt RJ (2007) Kinetic and Structural Studies on Interactions Between Heparin or Heparan Sulfate and Proteins of the Hedgehog Signaling Pathway. In: Biochemistry 46(13), pp. 3933–3941. doi:10.1021/bi6025424

Zheng H, Cooper DR, Porebski PJ, Shabalin IG, Handing KB, Minor W (2017) CheckMyMetal: A Macromolecular Metal-Binding Validation Tool. In: Acta Crystallographica. Section D, Structural Biology 73(Pt 3), pp. 223–233. doi:10.1107/S2059798317001061

Zielenkiewicz P, Rabczenko A (1984) Protein–Protein Recognition: Method for Finding Complementary Surfaces of Interacting Proteins. In: Journal of Theoretical Biology 111(1), pp. 17–30. doi:10.1016/S0022-5193(84)80193-9

Zielenkiewicz P, Rabczenko A (1985) Searching for Interacting Surfaces of Proteins—the Improved Method. In: Journal of Theoretical Biology 116(4), pp. 607–612. doi:10.1016/S0022-5193(85)80090-4

Zuckerman DM (2011) Equilibrium Sampling in Biomolecular Simulations. In: Annual Review of Biophysics 40, pp. 41–62. doi:10.1146/annurev-biophys-042910-155255

A Appendix

Epitopsy energetic model

This section provides the theoretical background for electrostatics correlation. To this end, the Hamiltonian of a point-charge in continuum electrostatics is formulated in a statistical mechanics framework. This section builds upon a prior formulation of the Epitopsy energetic model in Wilms 2013, section 4.2.1 "Electrostatic interaction energy model".

Let P be a protein of charge density $\rho(\vec{r})$ generating an electrostatic potential $\Phi(\vec{r})$ and L a ligand of N atoms with partial charges q_i and individual positions $\vec{r}_i$. To completely describe the position and orientation of L relative to P, six variables are required: three Cartesian coordinates for the vector between the geometric centers of P and L, and three angles $\phi \in [0, 2\pi)$, $\theta \in [0, \pi]$, $\psi \in [0, \pi]$ to describe the orientation of L. The coordinates of atoms in L after a rotation are given by the transformation $\vec{r}_i{}' = \hat{R}(\phi, \theta, \psi) \cdot (\vec{r}_i - \vec{r}_\mathrm{L}) + \vec{r}_\mathrm{L}$ with $\vec{r}_\mathrm{L} = \frac{1}{N} \sum_i^N \vec{r}_i$ the geometric center of L and $\hat{R}$ the rotation matrix operator.

Indistinguishable non-interacting particles such as L occupy available microstates $(\vec{r}, \phi, \theta, \psi)$ with Coulomb energies $U_E(\vec{r}, \phi, \theta, \psi)$ (Equation A.1) according to the Maxwell–Boltzmann distribution $e^{-\beta U_E(\vec{r}, \phi, \theta, \psi)}$ with $\beta = 1/k_B T$. The system partition function Z (Equation A.2) is formulated as the integral of the Maxwell–Boltzmann distribution over the complete Cartesian and angular space. The angles are sampled independently from each other, with probability density term $\frac{1}{2}\sin(\theta)$ to avoid oversampling at the poles (González 2010) and term $\frac{2}{\pi}\sin^2\left(\frac{\psi}{2}\right)$ (Kendall and Moran 1963, paragraphs 4.25–4.30 on pages 93–101) to avoid undersampling at large ψ (Miles 1965; Rummler 2002; Rühle et al. 2013).

Since the partition function cannot be evaluated, it is necessary to define two intermediate quantities: the density of states $g(\vec{r})$ (Equation A.3) for L in the system {P + water + ions} and the density of states $g_0(\vec{r})$ (Equation A.4) for L in pure water, where the electrostatic potential is zero everywhere. $g(\vec{r})$ and $g_0(\vec{r})$ can be interpreted as the concentration of L in the bound and unbound states respectively, and their ratio yields the equilibrium constant $K(\vec{r})$ (Equation A.5), from which the electric Gibbs energy of binding $\Delta G^{\mathrm{elec}}(\vec{r})$ is deduced using the relation in Equation A.6.

$$U_E(\vec{r},\phi,\theta,\psi) = \sum_i^N q_i \Phi\left(\hat{R}(\phi,\theta,\psi)\cdot(\vec{r}_i - \vec{r}_{\mathrm{L}}) + \vec{r}_{\mathrm{L}} - \vec{r}\right) \tag{A.1}$$

$$Z = \int_{\mathbb{R}^3} \mathrm{d}^3\mathrm{r} \int_{-\pi}^{+\pi} \mathrm{d}\phi \frac{1}{2\pi} \int_0^{\pi} \mathrm{d}\theta \frac{1}{2}\sin(\theta) \int_0^{\pi} \mathrm{d}\psi \frac{2}{\pi} \sin^2\left(\frac{\psi}{2}\right) e^{-\beta U_E(\vec{r},\phi,\theta,\psi)} \tag{A.2}$$

$$g(\vec{r}) = \frac{1}{Z} \int_{-\pi}^{+\pi} \mathrm{d}\phi \frac{1}{2\pi} \int_0^{\pi} \mathrm{d}\theta \frac{1}{2}\sin(\theta) \int_0^{\pi} \mathrm{d}\psi \frac{2}{\pi} \sin^2\left(\frac{\psi}{2}\right) e^{-\beta U_E(\vec{r},\phi,\theta,\psi)} \tag{A.3}$$

$$g_0(\vec{r}) = \frac{1}{Z} \int_{-\pi}^{+\pi} \mathrm{d}\phi \frac{1}{2\pi} \int_0^{\pi} \mathrm{d}\theta \frac{1}{2}\sin(\theta) \int_0^{\pi} \mathrm{d}\psi \frac{2}{\pi} \sin^2\left(\frac{\psi}{2}\right) e^{0} = \frac{1}{Z} \tag{A.4}$$

$$K(\vec{r}) = \frac{g(\vec{r})}{g_0(\vec{r})} = \frac{1}{2\pi^2} \int_{-\pi}^{+\pi} \mathrm{d}\phi \int_0^{\pi} \mathrm{d}\theta \sin(\theta) \int_0^{\pi} \mathrm{d}\psi \sin^2\left(\frac{\psi}{2}\right) e^{-\beta U_E(\vec{r},\phi,\theta,\psi)} \tag{A.5}$$

$$\Delta G^{\mathrm{elec}}(\vec{r}) = -k_B T \ln\left(K(\vec{r})\right) \tag{A.6}$$

In a discretized simulation box, Equations A.1–A.6 become A.7–A.12, with Ω a finite ensemble of rotational states ω that map to discretized angular coordinates sampled from (ϕ,θ,ψ), $\Phi_{l,m,n}$ the electrostatic potential at discretized Cartesian coordinates (l,m,n) and $q^{\omega}_{l,m,n}$ the ligand charges after rotation $\hat{R}_{\omega}$. These equations are reproduced from Wilms 2013, section 4.2.1, equations 4.1–4.7.

$$U^{\omega}_{l,m,n} = \Phi_{l,m,n} \cdot q^{\omega}_{l,m,n} \tag{A.7}$$

$$Z = \sum_{l,m,n} \sum_{\omega}^{\Omega} e^{-\beta U^{\omega}_{l,m,n}} \tag{A.8}$$

$$g_{l,m,n} = \frac{1}{Z} \sum_{\omega}^{\Omega} e^{-\beta U^{\omega}_{l,m,n}} \tag{A.9}$$

$$g_{0_{l,m,n}} = \frac{1}{Z} \sum_{\omega}^{\Omega} e^{0} = \frac{|\Omega|}{Z} \tag{A.10}$$

$$K_{l,m,n} = \frac{g_{l,m,n}}{g_{0_{l,m,n}}} = \frac{1}{|\Omega|} \sum_{\omega}^{\Omega} e^{-\beta U^{\omega}_{l,m,n}} \tag{A.11}$$

$$\Delta G^{\mathrm{elec}}_{l,m,n} = -k_B T \ln\left(K_{l,m,n}\right) \tag{A.12}$$

B Appendix

Sampling over $SO(3)$

This section provides the background on the problem of sampling an isotropic set of points of arbitrary size over $SO(3)$, the special orthogonal group of dimension 3. This group contains all rotations in $\mathbb{R}^3$ that preserve distances, orientation and origin. The core of the issue is to partition the polar (θ), azimuthal (ϕ) and spin (ψ) coordinates isotropically. An ensemble of isotropically distributed points in $SO(3)$ is typically visualized as a uniform distribution of points on the surface of a 4-dimensional unit sphere.

There is a mathematical connection between sampling over $SO(3)$ and over S^2, the surface of a 3-dimensional unit sphere. Over S^2, the spin angle ψ is degenerate, which reduces the problem to partitioning the angular coordinates ϕ,θ isotropically. This problem is known as the optimal packing of spheres on a sphere (Mackay et al. 1977). Trivial solutions are obtained from platonic solids inscribed in a sphere: their vertices uniformly tessellate the sphere surface (Lindemann et al. 2005).

Although there is no exact formula to sample an arbitrary number of points over S^2, many algorithms will find approximate solutions. Quasi-uniform grids are obtained by increasing the number of vertices on platonic solids (Williamson 1970; Sadourny 1972; Cullen 1974; Cullen and Hall 1979). Uniform longitude–latitude grids sample Euler angles uniformly and independently from each other, leading to irregularities and oversampling at the poles (Swinbank and James Purser 2006; González 2010) that typically require corrective methods (Williamson 1992). More recent algorithms include Lattmanian angle grids (Lattman 1972; Williams 1973), Saff–Kuijlaars grids (Saff and Kuijlaars 1997), Sukharev grids (Sukharev 1971) and layered Sukharev grids (Yershova and LaValle 2004), heuristic grids (Edmundson 1992), spherical spiral grids (Chukkapalli et al. 1999) and Fibonacci grids (Hannay and Nye 2004; Swinbank and James Purser 2006; González 2010).

Returning to the $SO(3)$ sampling problem, a couple of special cases can be addressed first. For a spherical probe, such as Cu^{2+}or Cl^-, all rotation angles are degenerate due to spherical symmetry and a single rotational state can be used. For a rod-like probe, such as $N{\equiv}C{-}O^-$ or $O{=}N^+{=}O$, the spin angle ψ is degenerate due to radial symmetry along the main axis, and uniform sampling over $SO(3)$ simplifies to a uniform sampling over S^2 (Figure B.1). For probes with no degenerate rotation axis, it is necessary to solve the $SO(3)$ sampling problem directly. Trivial solutions can be derived from molecular symmetry, where the symmetry operations of the T_d, O_h and I_h groups generate uniform samples over $SO(3)$ of size 12, 24, and 60 for chiral

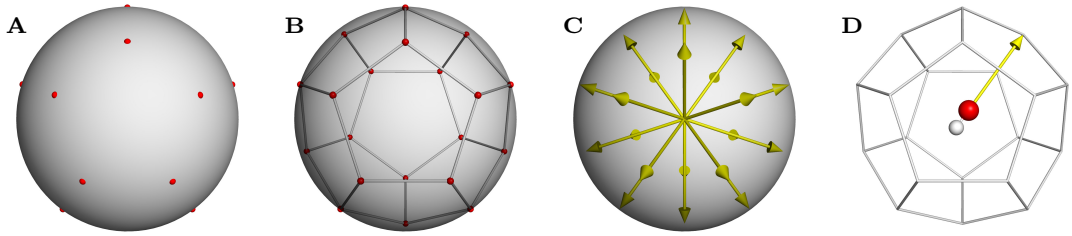


Figure B.1: Trivial solution to the $SO(3)$ sampling problem for rod-like probes and N=20. (A-B) Optimal sphere packing for N=20 (red dots) based on the regular dodecahedron (wireframe). (C) Optimal $SO(3)$ sampling for N=20. (D) Hydroxide ion (HO^-) with molecular axis collinear to one of the optimal vectors.

probes (orientation-preserving symmetry operations only). There is however no procedure to generate an arbitrary number of isotropically distributed points in $SO(3)$.

Multiple algorithms are available to generate approximately uniform samples over $SO(3)$ of arbitrary size. Several of them rely on the Lattman distance ℓ, which measures the rotation angle between two angular states (Lattman 1972), with $0 \leq \ell \leq \pi$. This distance can be visualized as the angle associated with the shortest arc connecting two points obtained from the application of two rotation matrices R_1 and R_2 on a vector $\vec{v}$. The limit case $\ell = 0$ corresponds to $R_1 = R_2$ and the limit case $\ell = \pi$ corresponds to a setup where $R_1\vec{v} = -R_2\vec{v}$.

In MolFit (Katchalski-Katzir et al. 1992), Euler angles (α, β, γ) (with $\alpha, \gamma \in [0, 2\pi], \beta \in [0, \pi]$) are sampled uniformly and independently with constant step. In FTDock (Gabb et al. 1997), an additional filtering step is carried out where pairs of angular states satisfying $\ell \leq t$ (with t a threshold) are considered degenerate and removed. In SOFTDOCK (Jiang and Kim 1991), Euler angles (α, β, γ) are generated such that each angular state has at least 4 neighbors within a threshold t. In Lattmanian grids, the angles are not sampled independently to limit oversampling at the poles (Lattman 1972). In the axis-angle method (Miles 1965), a rotation axis ϕ, θ is sampled over S^2 and the spin ψ is sampled from a non-uniform distribution with probability density $\sin^2(\psi/2)$. In Epitopsy, a rotation axis ϕ, θ is sampled over S^2 with the Fibonacci method (Hannay and Nye 2004; Swinbank and James Purser 2006; González 2010) and the spin ψ is sampled from a uniform distribution.

The uniformicity of a set of rotation angles sampled over $SO(3)$ can be assessed by examining the Lattman distance matrix of the set. As a first step, one can extract the distance to nearest neighbor (DNN) quantity u_i, which is the minimum of the off-diagonal elements of the matrix row-wise, i.e.

$$u_i = \min_{j \in J, j \neq i} \ell_{ij}, \; i \in J \tag{B.1}$$

with $J = 1, \ldots, N$ and N the size of the set. In a DNN plot, the points should concentrate on a single horizontal line if the sampling over $SO(3)$ is uniform. This criterion is not sufficient to conclude if a distribution is uniform, however if the criterion is not met, the distribution cannot be uniform.

DNN plots were produced for various implementations of the methods described previously, using a final set size of $N = 400$ whenever possible to facilitate comparison between methods (Figure B.2). Plots of the probability density of the DNN for the same methods are also provided for completeness (Figure B.3).

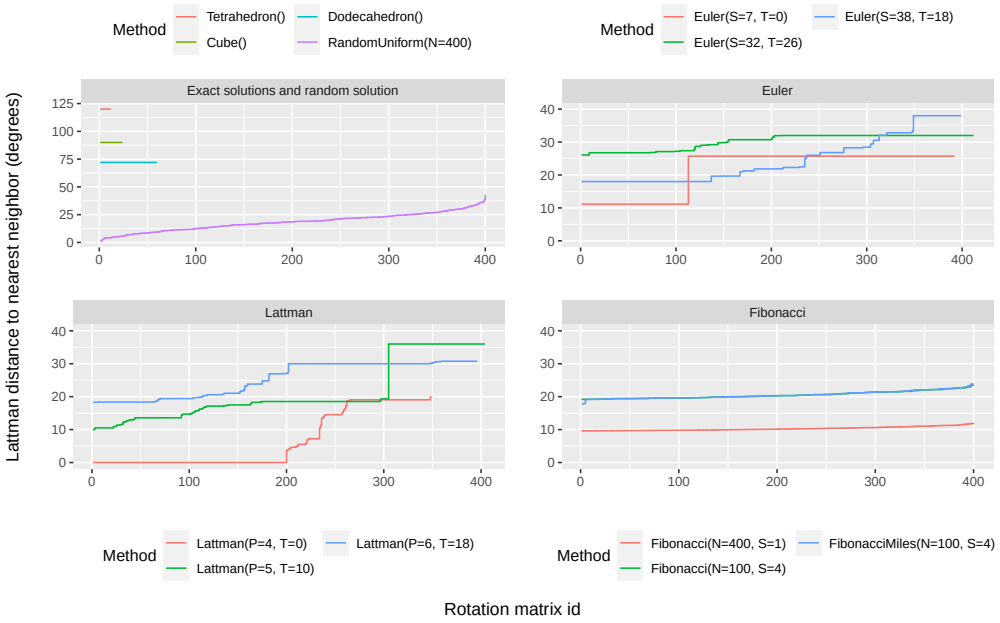


Figure B.2: DNN plots for multiple $SO(3)$ sampling methods. The same data is plotted as probability densities in Figure B.3. The legend provides the name and parameters of the Epitopsy methods used to generate the data.

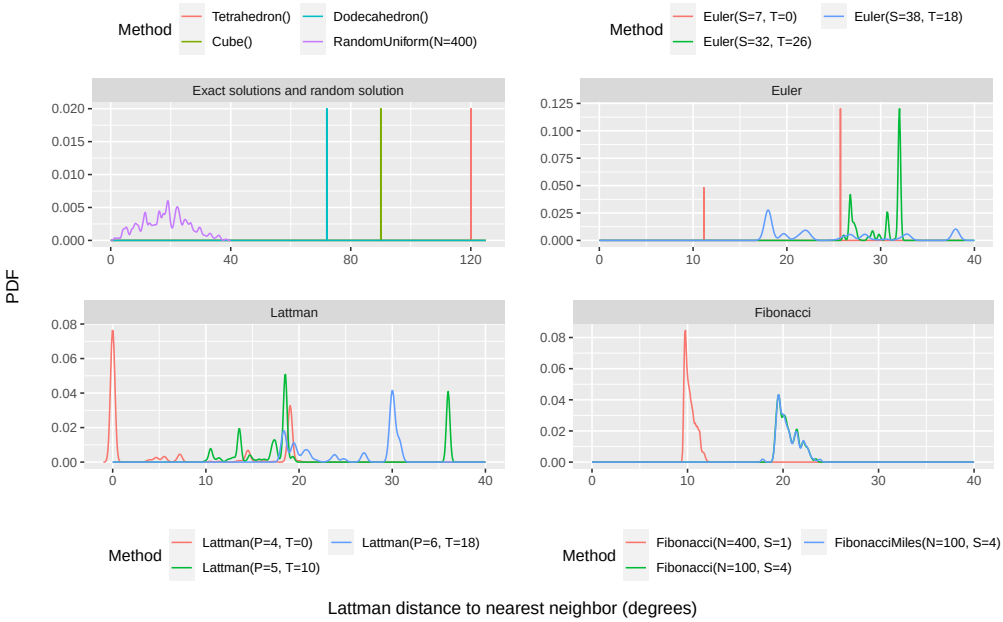


Figure B.3: DNN probability density for multiple sampling methods. A smoothing kernel was applied when necessary.

The following observations can be made:

- molecular symmetry method: only orientation-preserving symmetry operations are used (Figure B.2, top-left panel, red/green/blue curves). Uniform samples in $SO(3)$ are obtained, for which the DNN is a constant that depends on the sample size.
- random uniform method: angles (ϕ, θ, ψ) are sampled from a random uniform distribution (Figure B.2, top-left panel, purple curve). The high dynamic range in u_i is an indication that many angular states are clustered together ($\ell < 10°$) while peripheral angular states lie far from their nearest neighbors ($\ell \geq 25°$).
- Euler method: angles (α, β, γ) are sampled with a regular interval S (Figure B.2, top-right panel, red curve). The jump from $\ell = 11.1°$ to $\ell = 25.7°$ suggests a strong bias in the distribution of points, which is to be expected considering the method is prone to oversampling at the poles. Filtering out angular states that are within a threshold angle T yields a DNN that is more consistent with a uniform sampling (green and blue curves), albeit with numerous irregularities.
- Lattman: angles (ϕ, θ, ψ) are sampled with the Lattman method (Figure B.2, bottom-left panel, red curve) with P the number of samples for θ_2 (Lattman 1972). The distribution clearly shows a cluster of 200 degenerate rotational states ($\ell = 0°$). Filtering out angular states within a threshold angle T yields improved DNN distributions (green and blue curves) that are more consistent with a uniform sampling, but the Lattman method does not seem to be qualitatively better than the Euler method.
- Fibonacci method: N rotation axes (ϕ, θ) sampled from a Fibonacci sequence (Figure B.2, bottom-right panel, red curve) with S spin angles ψ (green and blue curves). This method is characterized by a smooth progression consistent with a locally uniform sampling, and an almost horizontal distribution. The spin plays a crucial role in increasing the spread of points over $SO(3)$: the distributions with spin are offset by 10° compared to the distribution without spin. There is no major difference between a uniform sampling of ψ and the non-uniform sampling of ψ proposed in Miles 1965.

Plotting the probability distribution of the rows of a Lattman distance matrix for a set of angular states, excluding the diagonal elements, offers a more quantitative view of the sampling quality. In the ideal case of the exact solutions obtained from symmetry groups, all rows in the Lattman distance matrix have the same histogram (Figure B.4). For other sampling methods, it would be desirable for the distributions to be homogeneous with one another and skewed towards π. In the case of the Euler, Fibonacci with spin and Lattman methods (Figure B.5), the distributions are not perfectly homogeneous, indicating non-uniform sampling over $SO(3)$. The Euler method produces the most homogeneous histograms (Figure B.5, left panel), however the DNN has already shown irregularities in the distribution. The Lattman method is highly inhomogeneous, with three curves clearly visible (Figure B.5, second panel), which indicates the angular states are forming clusters. The Fibonacci with spin method is also inhomogeneous, but is closer to the Euler method. This information, combined with the DNN plot, suggests the Fibonacci with spin method is preferable over the Lattman and Euler methods.

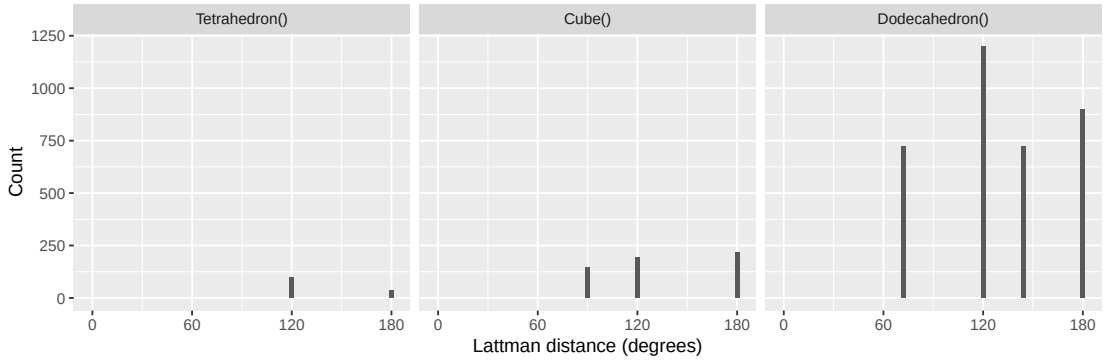


Figure B.4: Histograms of the Lattman distance matrices for exact solutions using molecular symmetry.

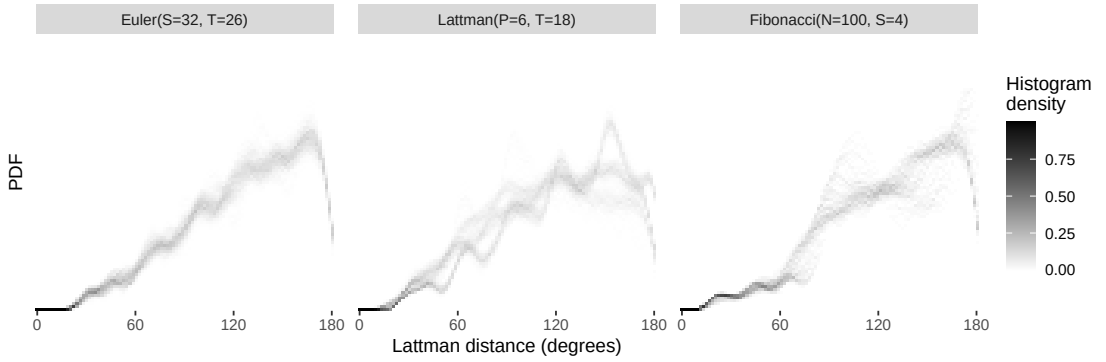


Figure B.5: Probability densities of the Lattman distance of sets generated using the Euler, Lattman and Fibonacci with spin methods, super-imposed for readability.

C Appendix

Partial charges derivation methods

The AMBER Hamiltonian defined in equation 1.1 relies on Coulomb's law for point charges. The partial charges q_i are obtained by fitting the potential of the point charge model outside the van der Waals surface of the molecule to the molecular electrostatic potential (MEP) from a QM calculation (Singh and Kollman 1984). Multiple fitting strategies were developed: least square fit with penalties to preserve electroneutrality (Cox and Williams 1981) and reproduce the experimentally determined dipole moment (Momany 1978), Levenberg–Marquardt nonlinear least square fit (Singh and Kollman 1984), the method of Lagrange multipliers (Chirlian and Francl 1987; Besler et al. 1990), and the restrained electrostatic potential (REsP) method (Bayly et al. 1993; Cornell et al. 1993).

As the MEP is sensitive to the molecule conformation (Williams 1990; Reynolds et al. 1992; Stouch and Williams 1992), especially in hydrocarbon chains (Stouch and Williams 1992; Bayly et al. 1993), it is beneficial to run the fitting procedure with the MEP of multiple conformers. The results can be combined using ensemble averages with iterative MD sampling (Basma et al. 2001) or Lagrange multipliers (Reynolds et al. 1992; Bayly et al. 1993; Cornell et al. 1993; Cieplak et al. 1995). The REsP method supports multi-conformational fitting and Lagrangians to average the charges of symmetry-equivalent atoms (Cieplak et al. 1995). It is also capable of constraining groups of atoms to a specific charge, which is necessary to handle the boundary conditions of monomers. For example, amino acids are processed as blocked tri-peptides, using acetic acid (abbreviated ACE) as the *N*-terminus capping group and *N*-methylamine (abbreviated NME) as the *C*-terminus capping group.

The REsP fitting method was applied to derive charges on amino acids and nucleic bases for the AMBER ff94 force field (Cieplak et al. 1995; Cornell et al. 1995), on saccharides for the GLYCAM_2000 (Woods and Chappelle 2000; Basma et al. 2001; Kirschner and Woods 2001) and GLYCAM_06 (Kirschner et al. 2008) force fields, and on the QQJ-096 building blocks used in Work 1 and Work 2. The following will briefly explore the two-stage REsP fitting procedure (Cieplak et al. 1995) used in the original AMBER force field (Cornell et al. 1995).

Starting from a generic amino acid AA, a tri-peptide ACE–AA–NME is formed (Figure C.1). The initial molecular structure is geometry optimized and the MEP is calculated at the HF/6-31* level of theory in the gas phase. This basis set is known to overestimate dipole moments, which is desirable for polar molecules since their dipole is usually larger in solution than in the gas phase (Stouch and Williams 1992; Bayly et al. 1993). For less polar molecules, this

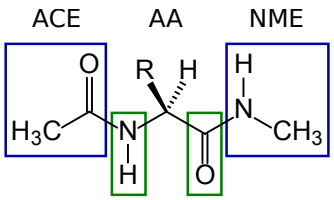


Figure C.1: Original REsP procedure for amino acids in the AMBER ff94 force field. The amino acid (AA) is capped at the *N*-terminus by an acetyl group (ACE) and at the *C*-terminus by an *N*-methylamine group (NME).

overestimation is problematic and has to be mitigated using a stronger hyperbolic penalty to decrease the magnitude of the charges (Woods and Chappelle 2000).

During the first REsP stage, several atoms are assigned fixed partial charges. The ACE and NME residues (blue boxes) are given standard charges that are identical for all amino acids to avoid introducing artificial dipoles on the protein backbone. These charges sum up to zero to preserve the formal charge on the amino acid. The backbone carbonyl group and amide nitrogen (green boxes) are given consensus charges that are identical for groups of amino acids sharing a chemically similar side-chain. Partial charges on the remaining atoms are allowed to vary during the fit, with intramolecular charge equivalence constraints if necessary[7] and intermolecular charge equivalence constraints when multiples molecules are used[8] .

During the second stage, all atoms get their charges fixed to those obtained in the first stage, except for symmetry-equivalent non-polar hydrogens, namely methyl hydrogens (R–CH_3) and freely rotatable methylene hydrogens (R–CH_2–R'), which are re-fitted with intramolecular charge equivalent constraints (Cieplak et al. 1995). When no such group exists, the second stage is unnecessary. The role of this stage is to avoid the formation of small, artificial dipole moments along freely rotatable hydrocarbon chains. This equivalencing is typically not conducted in the first stage to limit the number of Lagrangian constraints in a single fitting step; attempts to carry out both fits in a single step resulted in unrealistic dipole moments across bonds (Cornell et al. 1993; Bayly et al. 1993).

Here is a detailed protocol for lysine:

- *ab initio* geometry optimization and MEP calculation on two side-chain rotamers of lysine (Figure C.2**A**)
- REsP stage 1: fixed charges on the ACE, NME, amide and carbonyl groups (Figure C.2**B**, gray atoms), free fitting on aliphatic hydrogens (black atoms), fitting with intermolecular equivalence restraints (colored atoms) and an extra intramolecular equivalence restraint on the ammonium R–NH_3^+ group due to rotational symmetry
- REsP stage 2: charges are read from the previous stage (Figure C.2**C**, gray atoms), except for aliphatic methylene hydrogens (colored atoms) which are re-fitted with intra- and intermolecular equivalence restraints

[7] For example, C_3-equivalent ammonium hydrogens in lysine, C_2-equivalent aromatic carbons in phenylalanine and tyrosine, or C_2-equivalent guanidino R–NH_2 groups in arginine.

[8] When multiple conformers of the same molecule are used, identical atoms are constrained to be assigned the same partial charges. When different molecules are used, e.g. nucleic acids or *N*- and *C*-terminal amino acids, subsets of atoms on each molecule may be constrained to sum up to a specific charge (Cieplak et al. 1995).

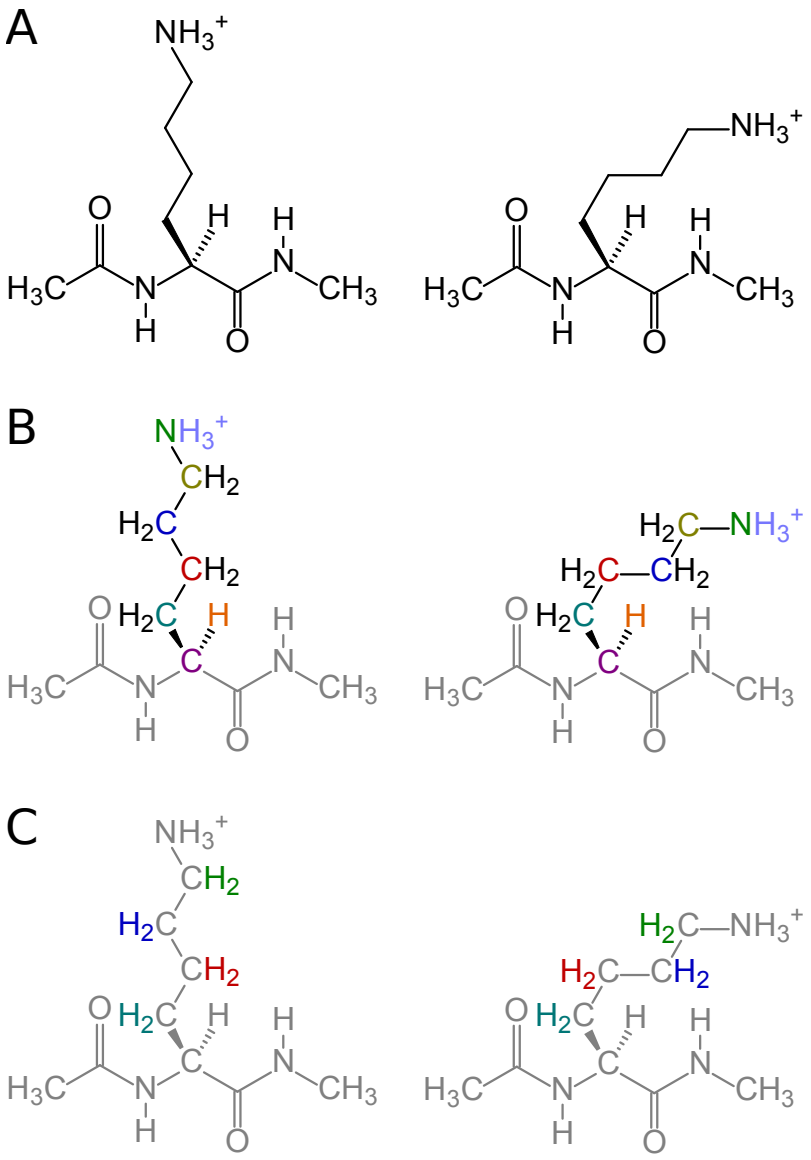


Figure C.2: Original REsP procedure for lysine in the AMBER ff94 force field. A) Optimized geometry of two conformers of lysine tri-peptide (ACE–LYS–NME), more specifically two rotamers differing by their Cα–Cβ dihedral angle (represented in the 2D sketch as a bent side-chain for simplicity). B) First stage of the REsP fitting procedure with atoms colored according to their restraints (gray: atoms constrained to fixed charges, black: free fitting, colored: fitting with intermolecular equivalence constraints and with an additional intramolecular equivalence constraint for the ammonium hydrogens). C) Second stage of the REsP fitting procedure with atom colored according to their restraints (gray: atoms constrained to fixed charges read from the previous stage, colored: fitting with intra- and intermolecular equivalence constraints).

Erklärung:

Hiermit erkläre ich, gemäß § 7 Abs. (2) d) + f) der Promotionsordnung der Fakultät für Biologie zur Erlangung des Dr. rer. nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient, bei der Abfassung der Dissertation nur die angegeben Hilfsmittel benutzt und alle wörtlich oder inhaltlich übernommenen Stellen als solche gekennzeichnet habe.

Stuttgart, den _____ _____

Unterschrift des/r Doktoranden/in

Erklärung:

Hiermit erkläre ich, gemäß § 7 Abs. (2) e) + g) der Promotionsordnung der Fakultät für Biologie zur Erlangung des Dr. rer. nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät/Fachbereich abgelehnt worden ist.

Stuttgart, den _____ _____

Unterschrift des/r Doktoranden/in

Erklärung:

Hiermit erkläre ich, gemäß § 6 Abs. (2) g) der Promotionsordnung der Fakultät für Biologie zur Erlangung der Dr. rer. nat., dass ich das Arbeitsgebiet, dem das Thema „Computational characterization of biomolecular interaction specificity“ zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Jean-Noël Grad befürworte und die Betreuung auch im Falle eines Weggangs, wenn nicht wichtige Gründe dem entgegenstehen, weiterführen werde.

_____

Name des Mitglieds der Universität Duisburg–Essen in Druckbuchstaben

Essen, den _____ _____

Unterschrift eines Mitglieds der Universität Duisburg–Essen

Kumulative Dissertation/Beteiligung an Veröffentlichungen

Kumulative Dissertation von Herrn Jean-Noël Grad

Autorenbeiträge

Titel der Publikation: Locating Large, Flexible Ligands on Proteins

Autoren: Jean-Noël Grad, Alba Gigante, Christoph Wilms, Jan Nikolaj Dybowski, Ludwig Ohl, Christian Ottmann, Carsten Schmuck, Daniel Hoffmann

Anteile:

- Konzept: 75%
- Durchführung der Experimente: NA%
- Datenanalyse: 100%
- Artenanalyse: NA%
- Statistische Analyse: 100%
- Manuskripterstellung: 50%
- Überarbeitung des Manuskripts: 50%

Unterschrift Doktorand/in

Unterschrift Betreuer/in

Kumulative Dissertation/Beteiligung an Veröffentlichungen

Kumulative Dissertation von Herrn Jean-Noël Grad

Autorenbeiträge

Titel der Publikation: A new class of supramolecular ligands stabilizes 14-3-3 protein–protein interactions by up to two orders of magnitude

Autoren: Alba Gigante, Jean-Noël Grad, Jeroen Briels, Maria Bartel, Daniel Hoffmann, Christian Ottmann, Carsten Schmuck

Anteile:

- Konzept: 10%
- Durchführung der Experimente: 0%
- Datenanalyse: 15%
- Artenanalyse: NA%
- Statistische Analyse: 15%
- Manuskripterstellung: 15%
- Überarbeitung des Manuskripts: 15%

Unterschrift Doktorand/in

Unterschrift Betreuer/in

The curriculum vitae is not included in the online version for data protection reasons.

Publications

- N Rafieiolhosseini, M Killa, T Neumann, N Tötsch, **JN Grad**, A Höing, T Dirksmeyer, J Niemeyer, C Ottmann, SK Knauer, M Giese, J Voskuhl and D Hoffmann **(2022)** Computational Model Predicts Protein Binding Sites of a Luminescent Ligand Equipped With Guanidiniocarbonyl-Pyrrole Groups. In: Beilstein Journal of Organic Chemistry 18, pp. 1322–1331. doi:10.3762/bjoc.18.137
- A Meiners, S Bäcker, I Hadrović, C Heid, C Beuck, YB Ruiz-Blanco, J Mieres-Perez, M Pörschke, **JN Grad**, C Vallet, D Hoffmann, P Bayer, E Sánchez-García, T Schrader and SK Knauer **(2021)** Specific Inhibition of the Survivin–CRM1 Interaction by Peptide-Modified Molecular Tweezers. In: Nature Communications 12(1), p. 1505. doi:10.1038/s41467-021-21753-9
- K Szuttor, F Weik, **JN Grad** and C Holm **(2021)** Modeling the Current Modulation of Bundled DNA Structures in Nanopores. In: The Journal of Chemical Physics 154(5), p. 054901. doi:10.1063/5.0038530
- H Anzt, F Bach, S Druskat, F Löffler, A Loewe, BY Renard, G Seemann, A Struck, E Achhammer, P Aggarwal, F Appel, M Bader, L Brusch, C Busse, G Chourdakis, PW Dabrowski, P Ebert, B Flemisch, S Friedl, B Fritzsch, MD Funk, V Gast, F Goth, **JN Grad**, J Hegewald, S Hermann, F Hohmann, S Janosch, D Kutra, J Linxweiler, T Muth, W Peters-Kottig, F Rack, FHC Raters, S Rave, G Reina, M Reißig, T Ropinski, J Schaarschmidt, H Seibold, JP Thiele, B Uekermann, S Unger and R Weeber **(2021)** An Environment for Sustainable Research Software in Germany and Beyond: Current State, Open Challenges, and Call for Action [version 2; peer review: 2 approved]. In: F1000Research 9(295). doi:10.12688/f1000research.23224.2
- D Smolin, N Tötsch, **JN Grad**, J Linders, F Kaschani, M Kaiser, M Kirsch, D Hoffmann and T Schrader **(2020)** Accelerated Trypsin Autolysis by Affinity Polymer Templates. In: RSC Advances 10(48), pp. 28711–28719. doi:10.1039/d0ra05827k
- A Gigante, **JN Grad**, J Briels, M Bartel, D Hoffmann, C Ottmann and C Schmuck **(2019)** A New Class of Supramolecular Ligands Stabilizes 14-3-3 Protein–Protein Interactions by Up to Two Orders of Magnitude. In: Chemical Communications (Cambridge, England) 55(1), pp. 111–114. doi:10.1039/c8cc07946c
- M Ehlers, **JN Grad**, S Mittal, D Bier, M Mertel, L Ohl, M Bartel, J Briels, M Heimann, C Ottmann, E Sanchez-Garcia, D Hoffmann and C Schmuck **(2018)** Rational Design, Binding Studies, and Crystal-Structure Evaluation of the First Ligand Targeting the Dimerization Interface of the 14-3-3ζ Adapter Protein. In: ChemBioChem: A European Journal of Chemical Biology 19(6), pp. 591–595. doi:10.1002/cbic.201700588
- **JN Grad**, A Gigante, C Wilms, JN Dybowski, L Ohl, C Ottmann, C Schmuck and D Hoffmann **(2018)** Locating Large, Flexible Ligands on Proteins. In: Journal of Chemical Information and Modeling 58(2), pp. 315–327. doi:10.1021/acs.jcim.7b00413
- S Pellegrini, **JN Grad**, T Bousquet and L Pélinski **(2011)** A Novel Multicomponent Reaction: Easy Access to Ferrocenyl (Alkylimino)-1,4-Dihydroquinolines. In: Tetrahedron Letters 52(15), pp. 1742–1744. doi:10.1016/j.tetlet.2011.01.144

Posters and talks

- **JN Grad**, R Weeber, M Holzer and C Holm **(2022)** Making Use of Established Codes Instead of Re-Inventing the Wheel: Using Lattice-Based Methods From waLBerla Into ESPResSo. In: SE22 – Software Engineering. (online event, Department of Computer Science, Humboldt-Universität zu Berlin, Germany, Feb. 21, 2022–Feb. 25, 2022). URL: https://pad.gwdg.de/XsfpS3qtSNOgSPcMKIEeCw?view#Making-use-of-established-codes-instead-of-re-inventing-the-wheel-Using-lattice-based-methods-from-waLBerla-into-ESPResSo
- **JN Grad**, R Weeber, F Weik, K Szuttor and C Holm **(2019)** ESPResSo: Experiences in Lowering the Entry Barrier for New Developers in a Particle-Based Simulation Package. In: deRSE19 – Conference for Research Software Engineers in Germany. (Albert Einstein Science Park, Potsdam, Germany, June 4, 2019–June 6, 2019). URL: https://derse19.uni-jena.de/derse19/talk/review/SGSKQL8N3FFV9W8TE8PTAFC3RVLWEYX3
- **JN Grad** and D Hoffmann **(2017)** Targeting Protein–Protein Interactions via Supramolecular Ligands – Mechanistic Insight by Molecular Modeling. In: Molecular Basis of Life, Abstracts. (Ruhr University Bochum, Bochum, Germany, Sept. 24, 2017–Sept. 27, 2017), p. 67. URL: https://gbm-online.de/tagungskalender-details/molecular-basis-of-life-2017-herbsttagung-der-gbm-25.html?file=files/gbm/tagungen/fall%20meetings/2017_MBoL_Abstracts.pdf#page=67
- **JN Grad** and D Hoffmann **(2017)** Molecular Dynamics Simulation Steady-State Prediction by Electrostatics — Applications in Supramolecular Chemistry. In: 11th Triennial Congress of the World Association of Theoretical and Computational Chemists, Book of Abstracts. (Gasteig Cultural Center, Munich, Germany, Aug. 27, 2017–Sept. 1, 2017), p. 1042. URL: http://www.watoc2017.com/files/WATOC17/Downloads/Book_of_Abstracts_final.pdf#page=1042
- **JN Grad** and D Hoffmann **(2017)** Molecular Dynamics Simulation Steady-State Prediction by Electrostatics – Applications in Supramolecular Chemistry. In: 2nd International Symposium: Supramolecular Chemistry on Proteins. (University of Duisburg-Essen, Essen, Germany, Sept. 20, 2017–Sept. 21, 2017), p. 14. URL: https://www.uni-due.de/imperia/md/content/crc1093/download/crc1093_international_symposium_2017_-poster_abstract_booklet-final.pdf#page=14
- **JN Grad** and D Hoffmann **(2016)** Computational Investigation of Multivalent Peptidic Dendrimers. In: CRC1093 and CRC765 Graduate Student Symposium: Protein–Ligand Interactions. (GHotel Hannover, Kleefeld, Germany, Aug. 31, 2016–Sept. 2, 2016), p. 31. URL: https://www.uni-due.de/imperia/md/images/crc1093/graduate-school/crc1093_and_crc765_graduate_student_symposium_2016_-_abstract_book_-_aug2016.pdf#page=31
- **JN Grad** and D Hoffmann **(2015)** Investigating Multiarmed Peptidic Dendrimers Using MD Simulations. In: CRC1093 International Symposium: Supramolecular Chemistry on Proteins. (University of Duisburg-Essen, Essen, Germany, Sept. 29, 2015–Sept. 30, 2015), p. 11. URL: https://www.uni-due.de/imperia/md/content/crc1093/download/crc1093_international_symposium_2015_-poster_abstract_booklet-sept19.pdf#page=11
- **JN Grad** and D Hoffmann **(2014)** Parameterizing Novel Residues for the AMBER99SB Forcefield. In: First CRC1093 Graduate Symposium: Supramolecular Chemistry on Proteins. (Schloss Gnadenthal, Kleve, Germany, Sept. 22, 2014–Sept. 24, 2014)