Faculty of Engineering and Sciences
Pontificia Universidad Javeriana-Cali

PhD Thesis

In Silico Design of Tweezer-like Molecules for the Specific Detection of Sucrose at Ultra-Low Concentrations

Gustavo Adolfo Lara-Cruz

Advisor: **Andrés Jaramillo-Botero, PhD.**

February 14, 2025

Copyright Statement

Gustavo Adolfo Lara-Cruz

Thesis submitted in partial fulfillment of the requirements for the degree of
Ph.D. in Engineering and Applied Sciences

Pontificia Universidad Javeriana-Cali
2025

Abstract

Measuring Sucrose (Suc) under real crop-field conditions is crucial for breeding plant species with enhanced resilience to biotic and abiotic stresses. Current quantification methods often disrupt plant's biochemical processes and lack the spatiotemporal resolution needed to monitor sucrose levels across tissues, hindering correlations between sucrose concentrations and metabolic responses under stress.

Tweezer-like molecules based on Diarylboronic Acid (DBA) offer a promising solution for selective sucrose quantification under *in vivo* conditions. These molecules reversibly react with carbohydrates, and their reaction products can be detected using techniques like Field-Effect Transistors (FETs), electrochemistry, or spectroscopy, making them suitable for portable on-field devices. While DBA tweezers have been used for glucose quantification, a systematic methodology for adapting them to sucrose detection is lacking.

In this work, we developed novel DBA tweezers for selective sucrose binding using an inverse design strategy supported by quantum chemistry calculations. Our approach aimed to: (i) elucidate sucrose reactivity with Arylboronic Acid (aBA) molecules in aqueous media, (ii) design DBA tweezers with optimized boronic group orientation for selective sucrose binding, and (iii) identify synthetically accessible DBA tweezers for sucrose quantification under physiological conditions. We also proposed a synthesis plan for one of the top-performing DBA tweezers, and outlined a methodology for its experimental implementation in Au surface with Self-Assemble Monolayer (SAM) coatings.

Our results show that aBA molecules react with sucrose's hydroxyl groups at the 1',3'- and 4,6-positions, forming stable $\mathrm{aBA}^{-1}-\mathrm{Suc}-\mathrm{aBA}^{-1}$ complexes with six-membered rings. Phenylboronic Acid $(\mathrm{PBA})^{-1}$ reacts spontaneously with sucrose at $\mathrm{pH} > \mathrm{p}K_{a,\mathrm{PBA}}$ on a timescale of minutes to hours. Using the complex $\mathrm{PBA}^{-1}(\mathrm{S})-\mathrm{Suc}-\mathrm{PBA}^{-1}(\mathrm{S})$ as scaffold, 1361 DBA tweezers for sucrose were designed.

A semi-automated computational workflow was developed to calculate the reaction-free energies for DBA tweezers with carbohydrates. Validation against experimental data for glu-

cose and galactose reactions demonstrated high accuracy (mean absolute error of 2.2 kcal/mol) and precision (Interquartile Range (IQR) of 2.4 kcal/mol). A Graph Based-Genetic Algorithm (GB-GA) multi-objective optimization completed the design process of the DBA tweezers. This approach identified DBA tweezers with optimal properties for sucrose quantification under physiological conditions, with selectivity for sucrose of around 1000 times over glucose.

The DBA tweezer 1,5-bis(3-boronophenyl)phenanthrene was used as a pilot molecule to explore the synthesis route for producing one of the top-performing DBA tweezers. Although the synthesis of the pilot molecule was not completed, the successful synthesis of 1,5-bis(3-methoxyphenyl)phenanthrene enabled us to optimize the reaction conditions for assembling the atomic backbone required for the pilot molecule.

To assess the performance of the DBA tweezer molecules with the phenanthrene linker, we conducted Molecular Dynamics (MD) simulations of the water-SAM-Au interface to estimated an optimal Au surface coverage, corresponding to 720 $Å^2$/tweezer. This surface coverage set a theoretical upper Limit of Detection (LOD) of 56.2 fM for a functionalized Au surface of 4.4 mm diameter. A FET device utilizing the Au-SAM surface as a gate in direct contact with the sample solution is expected to enable highly specific sucrose quantification in carbohydrate mixtures, at near 1:1 sucrose-to-carbohydrate concentration ratios, and constant chemical potential at the water-SAM interface—an advancement beyond the current state-of-the-art for synthetic sucrose receptors.

Sensors based on DBA tweezers tailored for sucrose quantification hold significant potential for continuous *in vivo* monitoring, providing insights into sucrose's role as a signaling molecule under stress. This work supports the development of real-time sucrose sensors for agricultural applications, advancing plant science and crop breeding strategies. The computational framework can be adapted for designing DBA tweezers for other carbohydrates, enabling simultaneous *in vivo* measurements and advancing our understanding of metabolic networks.

Keywords: Sucrose quantification; Tweezer-like diboronic acids; *In vivo* carbohydrate sensing; Computational inverse design; Metabolic stress responses; Real-time biosensing.

Resumen

La medición de sacarose en condiciones reales de cultivo es crucial para el desarrollo de especies vegetales con mayor resiliencia a los estreses bióticos y abióticos. Los métodos de cuantificación actuales a menudo alteran los procesos bioquímicos de las plantas y carecen de la resolución espaciotemporal necesaria para monitorear los niveles de sacarosa en diferentes tejidos, lo que dificulta las correlaciones entre las concentraciones de sacarosa y las respuestas metabólicas en situaciones de estrés.

Las moléculas tipo pinza basadas en ácidos diarilborónicos (ADB) ofrecen una solución prometedora para la cuantificación selectiva de la sacarosa en condiciones *in vivo*. Estas moléculas reaccionan de forma reversible con los carbohidratos y sus productos de reacción se pueden detectar utilizando técnicas como transistores de efecto de campo, electroquímica o espectroscopia, lo que las hace adecuadas para dispositivos portátiles de campo. Si bien las pinzas de ADB se han utilizado para la cuantificación de la glucosa, falta una metodología sistemática para adaptarlas a la detección de la sacarosa.

En este trabajo, desarrollamos nuevas pinzas de ADB para la unión selectiva de sacarosa utilizando una estrategia de diseño inverso respaldada por cálculos de química cuántica. Nuestro enfoque apuntó a: (i) dilucidar la reactividad de la sacarosa con ácidos arilborónicos (AAB) en medios acuosos, (ii) diseñar pinzas de ADB con orientación optimizada del grupo borónico para la unión selectiva con sacarosa, e (iii) identificar pinzas de ADB sintéticamente accesibles para la cuantificación de sacarosa en condiciones fisiológicas. También propusimos un plan de síntesis para una de las pinzas de ADB con mejores prestaciones, y una metodología para su implementación experimental en superficies de Au con recubrimientos de monocapa de autoensamblaje (MAE).

Nuestros resultados muestran que las moléculas de AAB reaccionan con los grupos hidroxilo de la sacarosa en las posiciones 1', 3'- y 4,6-, formando complejos estables AAB^{-1}-Suc-AAB^{-1} con anillos de seis miembros. El ácido fenilborónico (AFB^{-1}) reacciona espontáneamente con la sacarosa a pH $>$ p$K_{a,\mathrm{AFB}}$ en una escala de tiempo de minutos a horas. Uti-

lizando el complejo $AFB^{-1}(S)-Suc-AFB^{-1}(S)$ como plantilla, se diseñaron 1361 pinzas de ADB para sacarosa.

Se desarrolló un flujo de trabajo computacional semiautomatizado para calcular las energías libres de reacción para pinzas de ADB con carbohidratos. La validación con datos experimentales para reacciones de glucosa y galactosa demostró una alta exactitud (error absoluto medio de 2.2 kcal/mol) y precisión (rango intercuartílico de 2.4 kcal/mol). Una optimización multiobjetivo junto a un algoritmo genético completó el proceso de diseño de las pinzas de ADB. Este enfoque identificó pinzas con propiedades óptimas para la cuantificación de sacarosa en condiciones fisiológicas, con una selectividad para la sacarosa de alrededor de 1000 veces sobre la glucosa.

La pinza de ADB 1,5-bis(3-boronofenil)fenantreno se utilizó como molécula piloto para explorar la ruta de síntesis para producir una de las pinzas de ADB con mejores prestaciones. Aunque no se completó la síntesis de la molécula piloto, la síntesis exitosa de 1,5-bis(3-metoxifenil)fenantreno nos permitió optimizar las condiciones de reacción para ensamblar la estructura atómica requerida para la molécula piloto.

Para evaluar el rendimiento de las moléculas de pinza de ADB con el enlace de fenantreno, realizamos simulaciones de dinámica molecular de la interfaz agua-MAE-Au para estimar una cobertura superficial óptima de Au, correspondiente a 720 $Å^2$/tweezer. Esta cobertura superficial estableció un límite de detección teórico de 56.2 fM para una superficie de Au funcionalizada de 4.4 mm de diámetro. Se espera que un transistor de efecto de campo que utilice la superficie de Au-MAE como compuerta, en contacto directo con la solución de muestra, permita la cuantificación altamente específica de sacarosa en mezclas de carbohidratos en relaciones de concentración cercanas a 1:1 entre sacarosa y otros carbohidratos, y un potencial químico constante en la interfaz agua-MAE-Au, lo que representa un avance del estado del arte para los receptores sintéticos de sacarosa.

Lose sensores basadoes en pinzas de ADB para la cuantificación de la sacarosa tienen un potencial significativo para el monitoreo continuo *in vivo*, lo que proporciona información sobre el papel de la sacarosa como molécula de señalización bajo estrés. Este trabajo respalda el desarrollo de sensores de sacarosa en tiempo real para aplicaciones agrícolas, lo que hace avanzar la ciencia vegetal y las estrategias de mejoramiento de cultivos. El marco computacional se puede adaptar para diseñar pinzas de ADB para otros carbohidratos, lo que permite mediciones simultáneas *in vivo* y mejora nuestra comprensión de las redes metabólicas.

Palabras Clave: Cuantificación de sacarosa; Pinzas moleculares de ácidos diborónicos; Cuantificación *in vivo*; Diseño computacional inverso; Respuesta de estrés metabólico;

Biosensado en tiempo real.

Dedication

To my mother, Carmen, who has always supported me and taught me how some sacrifices are necessary to achieve what you truly dream of.

To my wife, Carolina, for her companionship, support, understanding, encouragement, and love in this long journey.

To my father, Luis Alfonso, for teaching me about resilience.

To my brothers, Luis Carlos and Ana Milena, for all their love and support.

To my family—Carlos, Liliana, Fabio, and Duffay—for always taking care of me and supporting me whenever I needed it.

To my grandmothers, Ana and Aura, and my grandfather, Alejandro. They are always in my thoughts.

To my mother-in-law, Victoria, for her loving support.

To my cats, Tango and Tutina, who have taught me about everyone needs their own space, and for helping me wake up very early to feed them. XD

Epigraph

"*Inmenso es quien puede vencer los retos que impone su ser*"
"*Convierte tu sed en mares de intenso ¡Amor por tus sueños!*"
-Elkin Ramirez, Sin Miedo al Dolor, Kraken.

Acknowledgements

I want to thank all the people who helped me complete this work:
To my advisor, Prof. Andrés Jaramillo-Botero, for his confidence in me and the freedom to pursue my research in solving the challenging problem of developing a method for sucrose quantification. I am also grateful for his generous support throughout my doctoral studies. His thought-provoking questions always guided me in the right direction during this research journey.

I would like to express my deepest appreciation to Prof. Stefan Grimme for his helpful and encouraging feedback on my research work and for accepting me into his research group as part of my international internship in Bonn, Germany. I am also grateful to Thomas Rose for being a generous co-author and a welcoming host during my time in Bonn, and to all the members of the AK Grimme group for the stimulating scientific discussions and the fun times playing table tennis.

I am extremely grateful to Prof. Agnieszka Nowak-Król for accepting me into her research group and for her generous support of my work in organic synthesis during my international internship in Würzburg, Germany. My sincere thanks also go to Klaudia Szkodzińska, who guided my retraining as a synthetic organic chemist and helped me plan the synthesis of the pilot molecule 1,5-bis(3-boronophenyl)-8-aminophenanthrene. I also thank all the members of the AK Nowak-Król group for their kindness, hospitality, and many wonderful memories.

I am deeply indebted to the administrative staff at iÓmicas, who always took care of and supported me throughout my doctoral studies.

To my colleagues in the doctoral program at Pontificia Universidad Javeriana-Cali—thank you for your friendship and support.

This work was partially funded by the "OMICAS program: Optimización Multiescala In-silico de Cultivos Agrícolas Sostenibles (Infraestructura y validación en Arroz y Caña de Azúcar)" Scientific Ecosystem belonging to the Colombia Científica Program, sponsored by The World Bank, The Ministry of Science, Technology and Innovation (MINCIENCIAS), ICETEX, the Colombian Ministry of Education and the Colombian Ministry of Commerce, Industry and Turism, under GRANT ID: FP44842-217-2018, OMICAS Award ID: 792-61187.

Publications

This PhD dissertation is based on the work contained in the following papers:

- G. A. Lara-Cruz and A. Jaramillo-Botero, "Molecular Level Sucrose Quantification: A Critical Review," Sensors, vol. 22, no. 23, p. 9511, Dec. 2022, doi: 10.3390/s22239511.
- G. A. Lara-Cruz, T. Rose, S. Grimme, and A. Jaramillo-Botero, "Reaction-Free Energies for Complexation of Carbohydrates by Tweezer Diboronic Acids," The Journal of Physical Chemistry B, vol. 128, no. 38, pp. 9213–9223, Sep. 2024, doi: 10.1021/acs.jpcb.4c04846.

Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

aBA Arylboronic Acid

ACF Autocorrelation Function

ALPB Analitycal Linearized Poisson-Boltzmann

AO Atomic Orbital

BOB Benzoxaborole

C-PCM Conductor-like Polarizable Continuum Model

CCSD(T) Coupled Cluster with Single and Double excitations and perturbative Triple excitations

CD Circular Dichroism

CE Conformer Ensemble

CENSO Commandline Energetic SOrting

CREST Conformer-Rotamer Ensemble Sampling Tool

DBA Diarylboronic Acid

DFA Density-Functional Approximation

DFT Density Functional Theory

ECFP6 Extended Connectivity Circular Fingerprints with radius 3

EDC·HCl N-(3 – Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride

FET Field-Effect Transistor

Fru Fructose

Gal Galactose

GB-GA Graph Based-Genetic Algorithm

GBSA Generalized Born model augmented with the hydrophobic Solvent Accessible Surface Area

GC-MS Gas Chromatography-Mass Spectra

GGA Generalized Gradient Approximation

Glc Glucose

INV Sucrose Invertase

IQR Interquartile Range

IRC Intrinsic Reaction Coordinate

LDA Local Density Approximation

LOD Limit of Detection

LSDA Local Spin Density Approximation

MBA Mercaptobenzoic acid

MD Molecular Dynamics

MEP Minimum Energy Path

MES 4-Morpholineethanesulfonic acid

MTD Metadynamics

NCI Non-Covalent Inteaction

NEB-CI Nudged Elastic Band-Climbing Image

NHS N–Hydroxysuccinimide

NMR Nuclear Magnetic Resonance

PBA Phenylboronic Acid

PBS Phosphate Buffer Solution

PCM Polarizable Continuum Medium

PES Potential Energy Surface

PT Potentiometric Titration

QCG Quantum Cluster Growing

RMSD Root Mean Square Displacement

SA Synthetic Accesibility

SAM Self-Assemble Monolayer

SASA Solvent-Accessible Surface Area

SIE Self-Interaction Error

SMARTS SMILES arbitrary target specification

SMD Universal Solvent Model

SMILES Simplified Molecular Input Line Entry System

SPGE Screen-Printed Graphene Electrode

Suc Sucrose

SuSy Sucrose Synthase

TD-DFT Time Dependent Density Functional Theory

TS Transition State

UDP Uracil-Diphosphate

WE Working Electrode

Chapter 1

Introduction and Background

1.1 Introduction

Global warming has led to an increase in the frequency of extreme climatic events such as droughts and floods, disruptions in seasonal cycles,[1,2] shifts in micro-climatic conditions within agricultural regions,[1,2] the spread of pathogens in crops,[3] and soil acidification.[4] These sterssors impact the normal development, growth, and yield of crops. As a result, developing agricultural crops that can withstand these stresses has become essential for ensuring food security and maintaining productivity.[2,3,5] Achieving this, requires advances in epigenetic and genetic plant improvement, which in turn requires a deep understanding of the biochemical processes underlying biotic and abiotic stress responses and their effects on gene expression under a particular set of conditions.[5,6]

Sucrose is the primary energy source in plants and plays a crucial role as a marker in initiating or halting various biochemical reactions. Produced through photosynthesis, Suc is transported via the phloem from source tissues to sink tissues,[7–9] where it participates in numerous metabolic processes related to growth, development, regulation, and signaling.[8] However, many aspects of sucrose's mechanisms, particularly its role as a signaling molecule, remain largely unknown.[7,9] Addressing this knowledge gap requires the development of specialized analytical protocols, methods, and tools. These tools must meet several operation constraints, including biocompatibility, high spatiotemporal resolution, and the ability to quantify ultra-low concentrations of sucrose under real crop field conditions.[10–12]

The most common methods for sucrose quantification do not accomplish all the aforementioned requirements. While analytical chemistry methods can detect ultra-low concentrations of sucrose, they require extensive parameter optimization. Additionally, these methods

lack high spatiotemporal resolution, are unsuitable for on-field measurements, and disrupt sucrose-related biochemical processes due to the necessary sample preparation.[12] Enzyme biosensors, though capable of selective detection under *in vivo* conditions, require precise control of pH, ion concentration, and temperature for accurate operation. Furthermore, the sucrose reactions on enzyme invertase biosensors are irreversible, potentially altering the plant's homeostatic equilibrium.[12]

Alternative approaches, including temple receptors,[13] aptamers,[14] molecularly imprinted polymers,[15,16] and tweezer-like DBA ligands,[17,18] have been less explored for sucrose quantification, leaving significant opportunities for the development of new methodologies. A more detailed review of the advantages and disadvantages, as well as the limits of detection and quantification for various sucrose detection methods, can be found in reference 12. However, despite the extensive literature about the quantification monosacharides with tweezer DBAs, there is no general procedure for extrapolating these advantages to the quantification of other carbohydrates. Finding a selective DBA for a specific carbohydrate is often serendipitous. To address this issue, we propose a holistic approach to designing molecular tweezers for the selective detection of sucrose. Our state-of-the-art computational chemistry workflow enables the calculation of reaction-free energies of sucrose and tweezer DBAs, and the systematic optimization of physicochemical parameters and synthetic accessibility of the later.

In this work, we tackle the challenge of the accurate and selective quantification of sucrose under *in vivo* conditions by the development of new tweezer-like DBA ligands tailored for selective reaction with sucrose. DBA ligands react reversibly with carbohydrates,[19] and the reaction products can be detected using field effect transistors,[20] electrochemistry,[21–23] spectroscopy,[24–27] or spectrophotometry[21,28–33] techniques, with high potential to be used in small portable devices for on-field measurements of sucrose.

This chapter continues by highlighting the importance of sucrose in plants and the challenges of its selective quantification in aqueous media (Section 1.1.1). Section 1.2 presents the background for this work, beginning with the chemical principles for designing tweezer DBA molecules and advancements in the selective quantification of glucose using DBA tweezers (Section 1.2.1), followed by the modular synthesis of DBA tweezers (Section 1.2.1.1) and the computational design of DBA tweezers for glucose (Section 1.2.1.2). Finally, Section 1.2.2 presents the theoretical background of *in silico* methods employed in this work for designing DBA tweezers for selective sucrose quantification under physiological conditions.

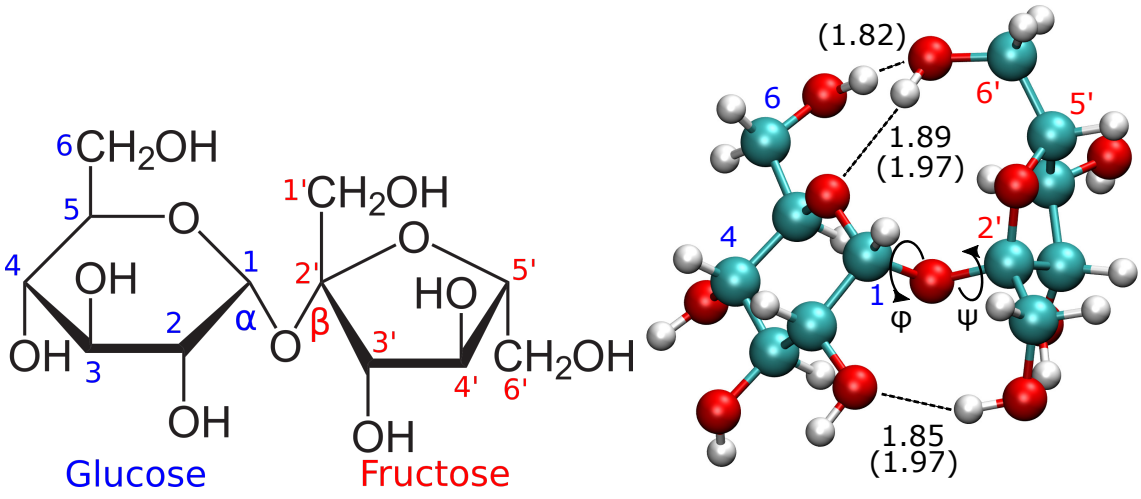


Figure 1.1: Haworth projection (left) and atomistic representations of sucrose (right). The atomistic representations show the hydrogen bond distances from the crystalline structure[34] and from ab initio calculations[35] (in parenthesis). Figure from reference 12.

1.1.1 The Role of Sucrose in Plants

Sucrose is composed of the monosaccharides glucose and fructose, linked by a glycosidic bond between their anomeric carbon atoms. This bond connects carbon atom 1 of glucose, in the α configuration, with carbon atom 2 of fructose, in the β configuration (as shown in Figure 1.1). This $1 \rightarrow 2$ glycosidic bond makes sucrose a non-reducing sugar with a high chemical stability, by locking the formations of acetal bonds needed for the formation of the aldehyde group in the mutarotation processes.[36] As a result, this bond inhibits spontaneous oxidation under physiological conditions and prevents the attachment of additional monosaccharide units.

The high chemical stability of sucrose under physiological conditions, combined with its relatively low viscosity in solution,[7] enables its efficient transport throughout the plant. Sucrose moves from source tissues, such as mature leaves where photosynthesis occurs, to sink tissues, which are either developing (e.g., meristematic tissues in young leaves and roots) or lack photosynthetic capacity (e.g., seeds, flowers, roots, and fruits).[7] In sink tissues, sucrose serves as an energy source, as a carbon atom source for building new tissue, and as a signaling molecule that regulates developmental and metabolic processes.[8] Sucrose is cleaved into its constituent units by two types of enzymes: Sucrose Invertase (INV) and Sucrose Synthase (SuSy). INV catalyzes an irreversible cleavage, producing D-glucose and D-fructose.[37] In contrast, SuSy catalyzes a reversible reaction, yielding Uracil-Diphosphate (UDP)-glucose and

D-fructose from sucrose and UDP, with the equilibrium favoring cleavage under physiological conditions.[7,38]

While the role of sucrose as a raw material in the early stages of metabolic and developmental processes in plants is relatively well understood, emerging evidence suggests that sucrose also functions as a key signaling molecule. It plays a crucial role in regulating carbohydrate metabolism, plant development through storage proteins, anthocyanin accumulation, floral induction, and the transport of nutrients to sink tissues.[8]

Currently, the functions of carbohydrates are elucidated using mutant plants with selectively modified genes that target specific processes related to each carbohydrate. Over time, the roles and significance of these carbohydrates are revealed through the phenotypic traits exhibited by the mutant plants. However, in the case of sucrose, many of its responses are also triggered by glucose,[8] complicating the understanding of sucrose's unique functions and mechanisms as a signaling molecule. Therefore, rapid quantification of sucrose on time scales of minutes to hours is crucial for several reasons: (i) sucrose has a high translocation rate, often exceeding 100 cm/h,[39] meaning that sucrose produced in source tissues quickly becomes available to sink tissues; and (ii) the rapid interconversion between sucrose and glucose makes it challenging to accurately estimate the specific amounts of sucrose responsible for eliciting plant responses to biotic or abiotic stresses. Consequently, there is a need for high-precision screening techniques to differentiate sucrose from other carbohydrates and to quantify its critical concentrations in specific metabolic responses.

In contrast to sucrose, the continuous quantification of glucose under *in vivo* conditions and in biological samples has seen significant progress, with several successful solutions developed. These range from biosensors capable of measuring glucose directly in plant tissues[40] or raw biological samples[41] to tweezer-like DBA molecules used for glucose continuous measurement in mice,[42] live cells,[43,44] zebrafish,[43] and plasma.[44]

Biosensors, however, are known to have limited lifetimes and require strict operational conditions. In comparison, tweezer-like DBA molecules are biocompatible, offer longer lifespans, and function under less rigorous conditions. Despite these advantages, there are currently no reports of sucrose-selective tweezer DBAs, nor is there a methodology to extend the demonstrated benefits of tweezer DBAs for *in vivo* glucose quantification to sucrose detection.

In the following sections, we present the theoretical background of this work, including the chemical principles underlying the design of tweezer-like DBA molecules for the selective quantification of carbohydrates in aqueous conditions.

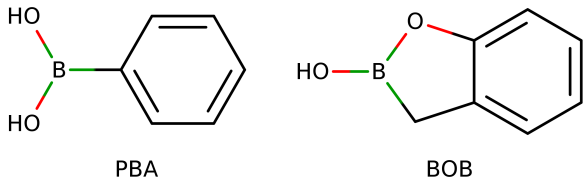


Figure 1.2: Phenylboronic Acid (PBA) and Benzoxaborole (BOB) molecules.

1.2 Background

1.2.1 Selective Detection and Quantification of Carbohydrates with Arylboronic Acids

Arylboronic Acids (aBAs), such as PBA and Benzoxaborole (BOB) (see Figure 1.2), have been widely used in the design of molecules for the separation and resolution of carbohydrate mixtures, particularly monosaccharides. These aBA molecules form covalent and reversible bonds with molecules containing diols at the 1,2- or 1,3- positions, resulting in a five- or six-membered borodiester ring, respectively.[45–47] Figure 1.3 illustrates the pH-dependent equilibrium reactions of aBAs with diols, leading to the formation of borodiester ring anions,[48] which represent the most stable species for an exoergic reaction between a diol and an aBA molecule. Furthermore, in a system involving a carbohydrate and an aBA molecule, the reactivity is highest at pH levels greater than or equal to the pK_a of the aBA and lower than the pK_a of the carbohydrate,[19,47] when the anionic form of the aBA molecule predominates.

The literature suggests that the optimal pH for the formation of the borodiester anion can be estimated as the average of the acid dissociation constants of the aBA (p$K_{a,\mathrm{aBA}}$) and the diol (p$K_{a,\mathrm{diol}}$) molecules:[49]

$$\mathrm{pH}_{\mathrm{optimal}} = \frac{\mathrm{p}K_{a,\mathrm{aBA}} + \mathrm{p}K_{a,\mathrm{diol}}}{2}. \tag{1.1}$$

According to Equation (1.1), PBA (p$K_a \approx 8$) binds optimally to monosaccharides such as glucose or fructose (p$K_a \approx 12$) at a pH of approximately 10. This makes PBA less suitable for carbohydrate detection under physiological conditions.[47] To address this limitation, various strategies have been developed to tailor the pK_a of aBAs molecules, enabling effective binding to diols at physiological pH. These strategies include introducing electron-withdrawing groups into the aromatic ring to increase the acidity of the boronic groups or incorporating

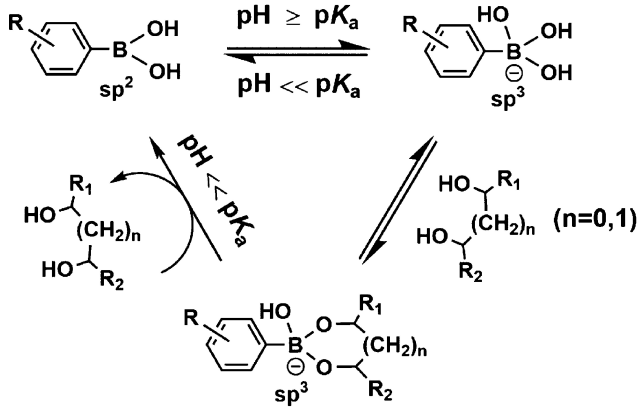


Figure 1.3: Reaction scheme for the complex formation of aBAs with diols. Figure taken from the reference 47.

substituents containing oxygen or nitrogen that donate electron pairs to the vacant p-orbital of the boron atom. Both approaches lower the p$K_{a,\mathrm{aBA}}$ of aBA molecules compared to PBA, as illustrated in Figure 1.4. These modifications also enhance the stability of the borodiester formed upon binding to diols.[47]

The stability and selectivity of the aryl boronic acids to react with a particular diol depend on steric effects,[48] orientation, relative position,[19,50] and acidity of hydroxyl groups, allowing the boronic acids to differentiate structurally similar saccharide molecules.[45] For example, PBA molecules form more stable products with fructose compared with glucose in aqueous media, with a equilibrium constant (K) of 4370 M^{-1} for PBA-Fructose (Fru) products,[1] compared to 110 M^{-1} for PBA-Glc products. This seems to be related with the higher acidity of the hydroxyl group on the anomeric carbon of fructose compared to that of glucose.[51] Moreover, PBA can form up to three boron ester bonds with the fructose hydroxyl groups at 2-, 3-, and 6-position.[52] However, in case of tweezer DBA molecules as presented in Figure 1.5, these molecules form more stable products with glucose rather than fructose.

Glucose, unlike fructose, has hydroxyl groups oriented correctly for binding with two aBA molecules.[18] The distance and positioning of boronic groups significantly influence the selectivity, reactivity, and types of products formed between tweezer-like DBA molecules and carbohydrates.[45,48,50,56] Figure 1.6 illustrates how the distance between the boronic groups ($\mathrm{R{-}B(OH)_3^-}$) modulates the types of products formed.[50] Even subtle changes in the position of the boronic group on the phenyl ring (*ortho-*, *meta-*, or *para-*) can alter the equilibrium

[1] In this document, we frequently use the acronyms for Glucose (Glc), Fructose (Fru), Galactose (Gal), and Sucrose (Suc) to refer to the carbohydrates present in the borodiester products formed with Arylboronic Acid (aBA) molecules.

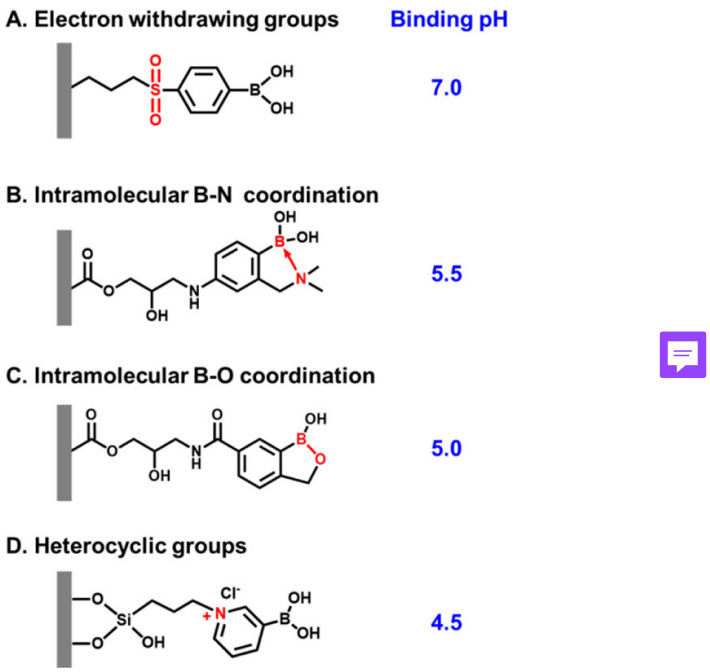


Figure 1.4: Strategies for the modulation of the pK_a in aBA molecules. Figure taken from reference 47.

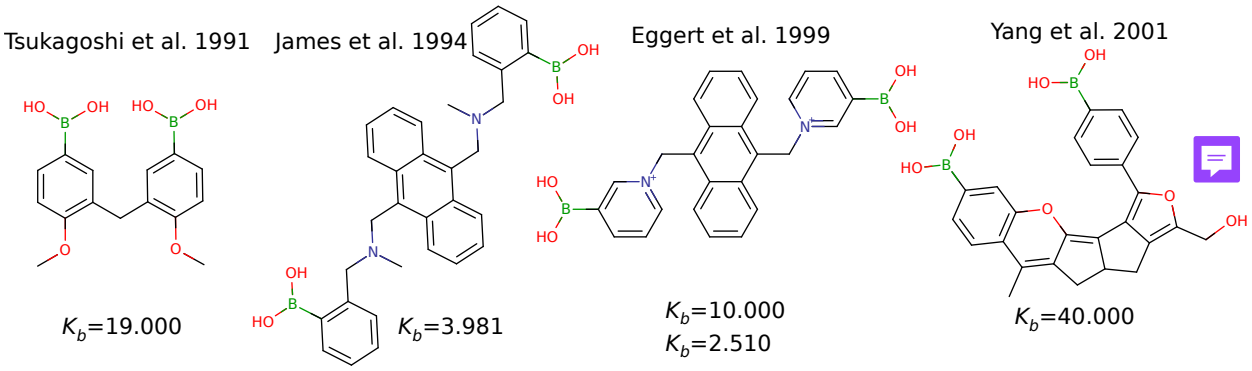


Figure 1.5: Examples of DBA molecules reported in the literature[17,53–55] with the corresponding equilibrium constants K_b for the tweezer DBA-Glc product.

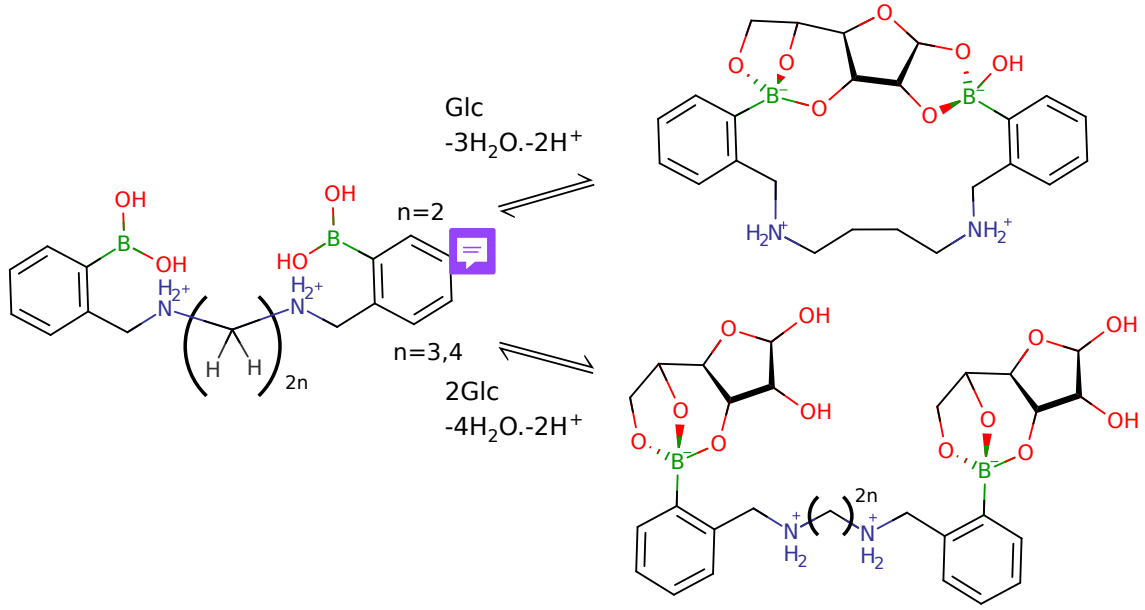


Figure 1.6: Reaction products of a DBA with glucose, performed under identical reaction conditions but different distances between the boronic acid groups by increasing the length of the alkyl chain $(-CH_2-)_{2n}$. A 1:1 reaction stoichiometry is observed for n=2, while a 1:2 stoichiometry is obtained for n=3 and n=4. Figure adapted from the reference 50.

constant K_b for the DBA-Glc product by an order of magnitude.[31]

The ability to bind two aBA molecules to glucose forms the chemical basis for the development of sensors for selective glucose quantification in aqueous media. For example, in the tweezer DBAs developed by James et al.,[54] glucose binding at two sites activates the tweezer's fluorescence by fixing the PBA's phenyl rings. Another type of sensor, based on glucose's dual binding to PBAs, includes highly selective sandwich-type sensors.[24–26,57] These sensors utilize surfaces and gold nanoparticles functionalized with PBA-S derivatives, taking advantage of the thiols self-assembling properties on gold surfaces. Measurements are conducted by observing changes in the Raman spectrum, amplified by surface plasmon resonance,[25,57] or by monitoring changes in the refractive index of the gold surface caused by the proximity of gold nanoparticles attached to glucose via PBA.[24,26] Using these refractive index measurements, glucose concentrations in the pM to nM range have been quantified.[24]

Another appealing property of certain tweezer DBAs is their straightforward synthesis via a modular approach, as detailed in the next section.

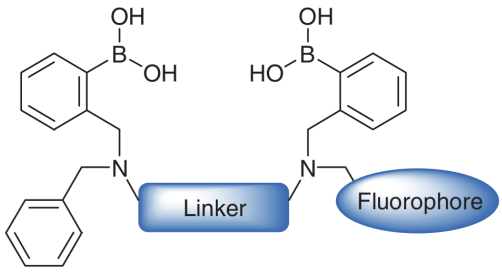


Figure 1.7: Modular design of ligands based in two molecules of PBA, a linker to tune the boronic groups separation, and a fluorophore molecule. Figure taken from the reference 19.

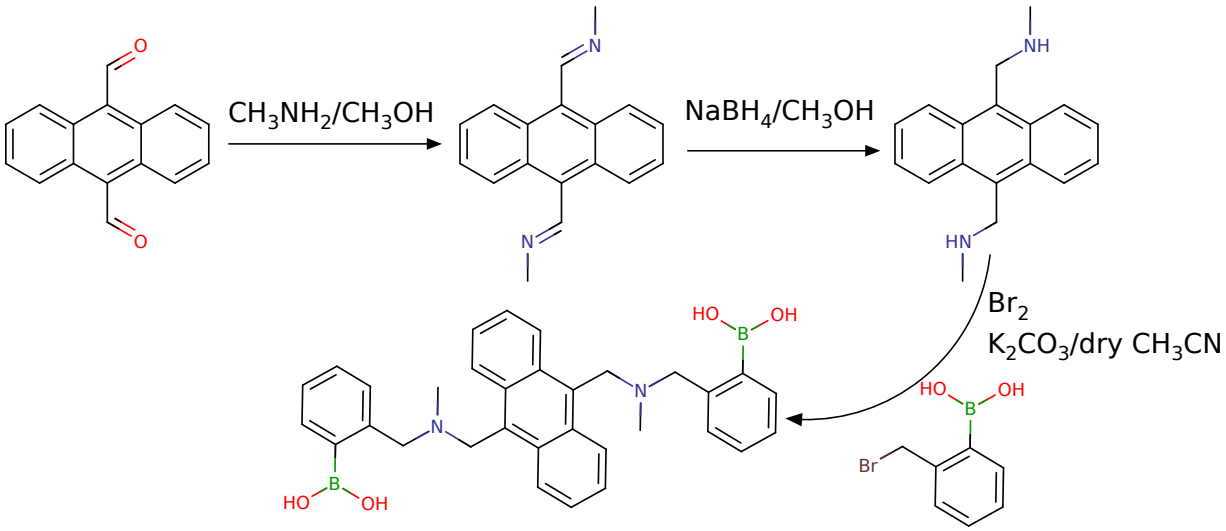


Figure 1.8: Synthesis of a modular tweezer ligand based on PBA. Figure reproduced from the reference 59.

1.2.1.1 Modular Synthesis of Tweezer Diboronic Acids

Some tweezer DBA for carbohydrates sensor have been rationalized as a modular ensembles of molecules (see Figure 1.7) in which the separation of the boronic groups is modulated by a linker molecule of a desired length of carbon atoms.[19,50] A fluorophore molecule, which is used to monitor the binding of sucrose to the phenyl boronic groups, is bonded to a tertiary amino group, and this allows the easy modification of the fluorophore molecule for its tuning to convenience.[19,56,58,59]

Figure 1.8 presents the synthesis steps for a ligand-based on the modular concept of Figure 1.7.[59] This modular ligand has the advantage of three steps for its synthesis starting from the commercially available compound 9,10-dialdehyde anthracene, where the anthracene unit is used as linker and fluorophore.

Although the concept of modular ligands based on aBAs offers an easy route for substituting linker and fluorophore units, the distance modulation of boronic groups using a six-carbon alkyl linker demonstrates only modest selectivity—1.2 times—for glucose over fructose.[19] The flexibility of most DBA tweezers synthesized via a modular approach may hinder improvements in selectivity ratios and limit the application of tweezer DBAs for the selective quantification of carbohydrates other than glucose.

To the best of our knowledge, the tweezer DBA designed by Yang et al.[17] is the only example that does not follow the modular design approach depicted in Figure 1.7. Instead, the authors employed an inverse design strategy, guided by computational chemistry calculations, to achieve optimal complementarity between the tweezer DBA and glucose. The design process and synthesis of Yang's DBA tweezer are detailed in the next section.

1.2.1.2 Computational Aided Design of Tweezer Diboronic Acids

Yang et al.[17] successfully designed a tweezer-like DBA featuring a dual glucose-binding site (see Figure 1.5), which demonstrated a binding constant of K_b = 40,000 M^{-1}—400 times greater than that for fructose or galactose. Their design began by identifying the optimal conformation for binding two PBA molecules to glucose, which was determined through molecular geometry optimization with quantum chemistry methods. The most stable geometric configuration was found when binding occurred at the hydroxyl groups located at positions 1,2 and 4,6 of glucose.[17] Using this approach, Yang and coworkers designed a PBA−Glc−PBA scaffold with optimally positioned PBA moieties for glucose binding, which served as the basis for developing a DBA tweezer. To complete the tweezer design, a carbon-based linker molecule was generated to connect the two PBA molecules, ensuring they remained fixed in their optimal positions within a single ligand structure. This linker fragment was created using the program CAVEAT, which employs a library of carbon-based molecules that can be merged according to specific chemical bonding rules. The fragment that best matched the size and orientation of the binding vectors, defined collinearly with the C−H bonds in the PBA aromatic ring, was selected. To facilitate synthesis, modifications were made to the carbon skeleton generated by CAVEAT.[17]

The high selectivity for glucose over fructose for the Yang's DBA tweezer[17] was achieved from the precise positioning of the boronic groups for the binding with the hydroxyl groups at 1,2- and 4,6-postions of glucose, and the rigidity of the organic backbone with aromatic rings that holds the boronic groups at its optimal positions.[17] In turn, an extensive synthesis procedure is needed for the organic backbone as presented in Figure 1.9. Probably, as con-

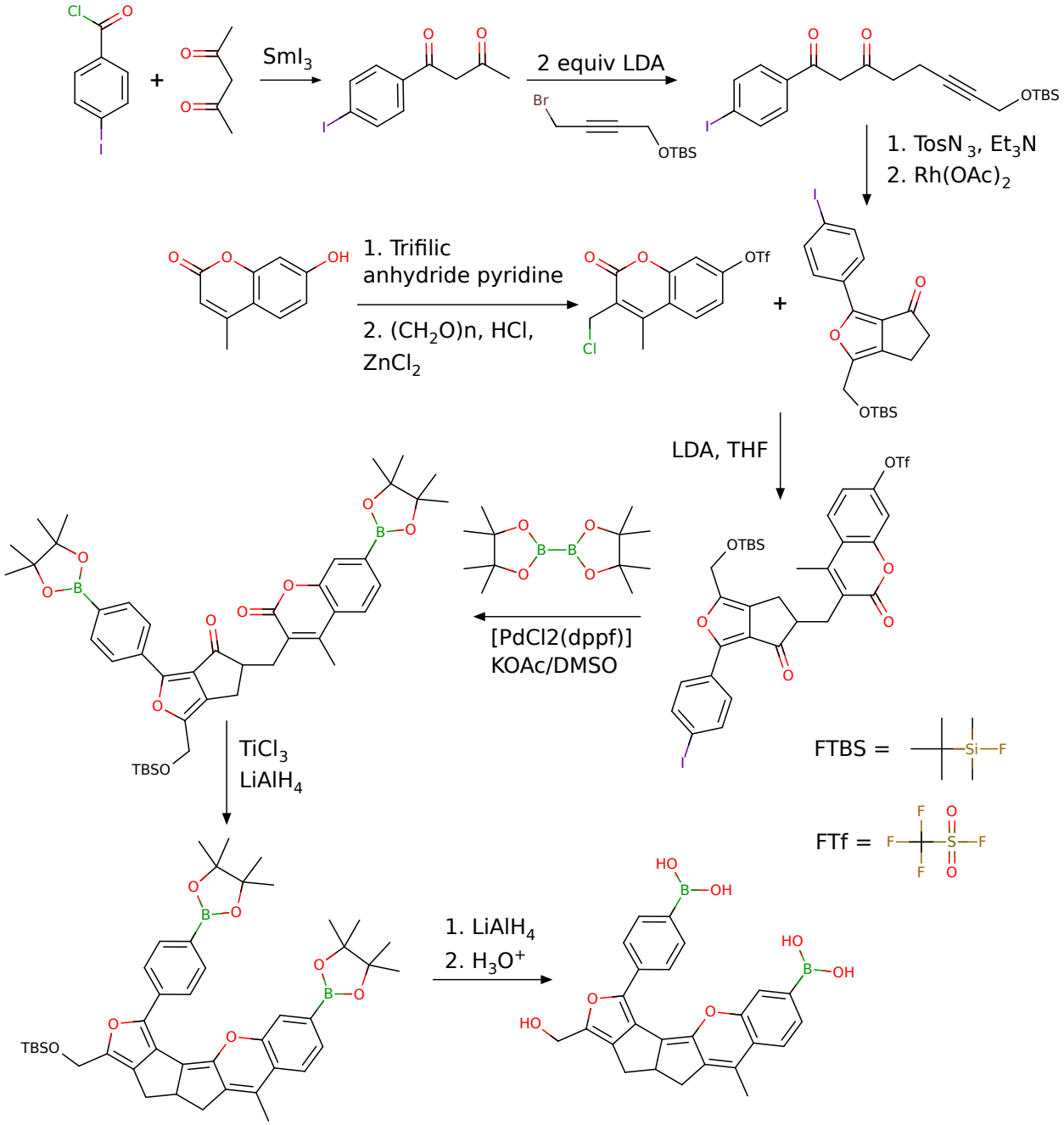


Figure 1.9: Reaction steps for the synthesis of the DBA tweezer reported by Yang et al.[17] Figure reproduced from the reference 17.

sequence of a lack of a metric for the synthesis complexity of the DBA tweezer during the desing process.

Although aryl boronic acids are extensively studied, mainly, for the detection of fructose and glucose, there are few reports on their application for sucrose detection.[20,30,31] Most studies treat sucrose as an interferent due to its low affinity for PBA-based molecules, and the boronic acid groups are not specifically oriented for binding to sucrose. However, Zhang et al.[31] reported a tweezer-like DBA molecule that, while selective for lactose ($K_b = 35\,\mathrm{M}^{-1}$), enabled the discrimination of disaccharides, including sucrose, within a mixture through linear discriminant analysis of the fluorescence spectrum in the millimolar concentrations.

The *in silico* design approach introduced by Yang and coworkers[17] in the early 2000s presented an interesting methodology for designing selective DBA tweezers for target carbohydrates through an inverse design strategy. However, since then, no major novel architectures of tweezer DBAs have been introduced. Recent studies on the selective quantification of glucose under *in vivo* conditions[43,44] primarily rely on modifications of the tweezer developed by James et al.[54] (see Figure 1.5).

Recent advancements in cost-efficient (semiempirical) quantum chemistry methods have enabled accurate calculations of inter-, intra-, and reactive interactions in systems with hundreds of atoms, such as tweezer DBAs interacting with carbohydrates. These developments have significantly enhanced computer-guided molecular design by expanding the availability of data on reactants, transition states, and reactions, consolidated in extensive databases.[60–65] Additionally, algorithms with well-defined chemical bonding rules have been developed for generating *de novo* molecules, employing approaches such as exhaustive database searches[60,61] or stochastic techniques like evolutionary algorithms,[62,66] among others. All these advancements have enhanced the efficiency of screening candidate molecules tailored for specific purposes, enabling the exploration of broader chemical spaces to discover molecules with better synthetic accessibility, and facilitating their subsequent experimental implementation.

The following sections provide the theoretical background of the *in silico* simulation methods used in this work for the computational design and screening of tweezer-like DBA molecules.

1.2.2 *In Silico* Simulations Methods

In silico simulation methods involve the use of computational techniques to: (i) model and predict molecular properties based on the microscopic description of electrons interacting with atomic nuclei as governed by quantum mechanical laws; (ii) calculate macroscopic properties

of matter from thermodynamic averages of microscopic properties derived from first-principle methods; and (iii) aid in the design of molecules with desired properties using computed thermodynamic characteristics. The following sections provide an overview of the theory supporting the *in silico* methods employed in this work.

1.2.2.1 The Potential Energy Surface

Matter at the molecular scale can be described by the quantum mechanics, therefore, the characteristics and properties of a molecular system can be calculated by solving the Schrödinger equation:

$$(\hat{T}_N + \hat{V}_{NN} + \hat{T}_e + \hat{V}_{ee} + \hat{V}_{Ne})\Psi_{mol}(R_N, r_e) = E_{mol}\Psi_{mol}(R_N, r_e) \tag{1.2}$$

where $\Psi(R_N, r_e)$ is the wave function for the quantum state of the molecular system that depends from the spatial coordinates for the atomic nuclei R_N and the electrons r_e; E_{mol} is the energy for the quantum molecular state; $\hat{T}_N$ and $\hat{T}_e$ are the kinetic energy operators for the nuclei and electrons, respectively; the nuclear and electronic repulsion energies are represented by the operators $\hat{V}_{NN}$ and $\hat{V}_{ee}$, respectively, and $\hat{V}_{Ne}$ is the operator for the nuclei-electron interaction energy.[67]

The solution of equation 1.2 for molecular systems, traditionally, has been performed by the use of the Born-Oppenheimer approximation in which the movement of the electrons are considered almost independent from the nuclei movement, allowing the decoupling of the nuclear (R_N) and electron (r_e) spatial coordinates. This approximation is based on the fact that the nuclei are around 1,800 times heavier than the electron.[67–69] In this way, the quantum molecular state $\Psi_{mol}(R_N, r_e)$ is written as a product function for the quantum states of the nuclei θ and the electrons ϕ:

$$\Psi_{mol} = \theta_N \, \phi_e \tag{1.3}$$

The molecular electronic energy E_e, for a fixed nuclear configuration R_N, is calculated from the solution of the electronic Schrödinger equation:

$$(\hat{T}_e + \hat{V}_{ee} + \hat{V}_{Ne}) \, \phi_e(r_e|R_N) = E_e \, \phi_e(r_e|R_N) \tag{1.4}$$

where $\phi(r_e|R_N)$ is the electronic state wave function for the nuclear configuration R_N.

An approximate solution for the molecular Schrödinger equation can be calculated from

the substitution of the equations (1.3) and (1.4) in (1.2):

$$(\hat{T}_N + \underbrace{\hat{V}_{NN} + E_e}_{U_{pot}})\,\theta_N(R_N)\phi_e(r_e|R_N) = E_{mol}\,\theta_N(R_N)\phi_e(r_e|R_N) \tag{1.5}$$

where U_{pot} is denotes the Potential Energy Surface (PES) for the nuclei movement. As the Equation 1.4 is a parametric function of the nuclei configuration, the calculation of E_n at all possible configurations R_N allows the calculation of the complete PES for the molecule. This is knows as the adiabatic PESs, because it is assumed that the electronic quantum state $\phi_e(r_e|R_N)$ does not change for a fixed R_N (this is valid only in some molecular configurations).

Based on the Born-Oppenheimer approximation, most of the algorithms used in computational chemistry have been developed to calculate and study the electronic structure of molecules, and from these electronic calculations it is possible to obtain the vibrational energy, dipole moments, perform analysis of reactivity, among others. The Born-Oppenheimer approximation is also important for studies on the dynamics of chemical reactions, since it provides the PES to describe the forces that govern the motion of nuclei during a reaction.[69–71] Once you have a PES for the reaction, it is possible to study the temporal evolution of the nuclei of the reactants from classical or quantum mechanics.

1.2.2.2 The Density Functional Approximation

The electronic wave function $\phi_e(r_e|R_N)$ for a molecule with $n-$electrons, is a high dimensional function that depends on $3n-$spatial coordinates and $n-$coordinates for the electronic spin. Hohenberg and Kohn[72] demonstrated that for a molecule in its ground state ψ_0, the energy and properties are univocally related with the three-dimensional electronic density of the ground state $\rho_0(x, y, z)$.[69] This is know as the Hohenberg and Kohn theorem, and motivated the development of the Density Functional Theory (DFT) in which the electronic energy of the ground state E_0 is considered as a functional of the electronic density ρ_0 ($E_0 = E[\rho_0]$).[69] From Equation 1.2, a functional relation is obtained from the expected value of the electronic energy:

$$E_0 = E_v[\rho_0] = \bar{T}[\rho_0] + \bar{V}_{Ne}[\rho_0] + \bar{V}_{ee}[\rho_0], \tag{1.6}$$

where:

$$\bar{V}_{Ne}[\rho_0] = \left\langle \psi_0 \middle| \sum_{n=1}^{n} \upsilon(\mathbf{r}_i) \middle| \psi_0 \right\rangle = \int \rho_0(\mathbf{r})\upsilon(\mathbf{r})d\mathbf{r}, \tag{1.7}$$

$\upsilon(\mathbf{r}_i) = \sum_\alpha Z_\alpha/\mathbf{r}_{i\alpha}$ is know as the external potential because it comes from the nuclei-electron interaction, which is external from the electrons system.[69] This external potential is determined once the density ρ_0 is know. Replacing Equation 1.7 into 1.6:

$$E_0 = E_\upsilon[\rho_0] = \int \rho_0(\mathbf{r})\upsilon(\mathbf{r})d\mathbf{r} + \bar{T}[\rho_0] + \bar{V}_{ee}[\rho_0] = \int \rho_0(\mathbf{r})\upsilon(\mathbf{r})d\mathbf{r} + F[\rho_0], \tag{1.8}$$

where $\bar{T}[\rho_0]$ and $\bar{V}_{ee}[\rho_0]$ are independent of the external potential. Equation 1.8 reduces the dimensionality for the calculation of the electronic energy but does not allow any practical calculation in chemistry because the functional relation of F with ρ_0 is unknown.[69]

Kohn and Sham[73] set it as a reference an ideal system of n not interacting electron gas for the calculation of the electronic energy and related properties of the molecule on its ground state with the DFT formalism. The difference between the kinetic energy average of the real ($\bar{T}$) and reference ($\bar{T}_s$) system is defined as:[69]

$$\Delta\bar{T}[\rho] \equiv \bar{T}[\rho] - \bar{T}_s[\rho], \tag{1.9}$$

and the difference of the electrostatic interaction energy between the real ($\bar{V}_{ee}$) and reference system is given by the equation:[69]

$$\Delta\bar{V}_{ee} \equiv \bar{V}_{ee}[\rho] - \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2. \tag{1.10}$$

From Equations 1.10 and 1.9, Equation 1.8 is rewritten as:

$$\begin{aligned} E_0 = E_\upsilon[\rho] &= \int \rho(\mathbf{r})\upsilon d\mathbf{r} + \bar{T}_s[\rho] + \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + \Delta\bar{T}[\rho] + \Delta\bar{V}_{ee}[\rho] \\ &= \int \rho(\mathbf{r})\upsilon d\mathbf{r} + \bar{T}_s[\rho] + \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_{ce}[\rho], \end{aligned} \tag{1.11}$$

where ρ is the density of the reference system, $E_{ci}[\rho] = \Delta\bar{T}[\rho] + \Delta\bar{V}_{ee}[\rho]$ is the functional of correlation and exchange, and its exact expression is unknown. There are several approximations for the $E_{ci}[\rho]$ functional as the Local Density Approximation (LDA), Local Spin Density Approximation (LSDA), Generalized Gradient Approximation (GGA), or hybrid approximations.[69]

Once selected a functional approximation for $E_{ci}[\rho]$, the calculation of the electronic molecular energy $E_\upsilon[\rho]$ is performed by the use of the variational theorem of Hohenberg and Kohn,[72] where the functional energy $E_\upsilon[\rho_{pr}]$ for a probe density $\rho_{pr}(\mathbf{r})$ subject to the

restrictions $\int \rho_{pr}(\mathbf{r})d\mathbf{r} = n$ and $\rho_{pr}(\mathbf{r}) \geq 0$, is always improved by the real density ρ_0.[69] The variational approximation for the electronic density ρ is equivalent to the variational approximation of the orbitals, and for the DFT energy calculations the next equation is used to calculate the electronic density for:

$$\rho = \rho_s = \sum_{i=1}^{n} |\Theta_i^{KS}|^2 \tag{1.12}$$

where Θ_i^{KS} are the Kohn-Sham orbitals, and the Θ_i^{KS} orbitals that minimize the Equation 1.11, are also solution for the equation:

$$\left[-\frac{1}{2}\nabla_1^2 - \sum_{\alpha} \frac{Z_\alpha}{r_{1\alpha}} + \int \frac{\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + \upsilon_{ce} \right] \Theta_i^{KS}(1) = \varepsilon_i^{KS} \Theta_i^{KS}(1) \tag{1.13}$$

where $\upsilon_{ce}(\mathbf{r})$ is know as the correlation-exchange potential defined by:[69]

$$\upsilon_{ce}(\mathbf{r}) \equiv \frac{\delta E_{ci}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \tag{1.14}$$

The use of the DFT calculations is not limited to the ground state of a molecule. The calculation of molecular energies of excited states and related properties as excitation energies, dipole transitions, absorption spectra are possible with the use of the Time Dependent Density Functional Theory (TD-DFT), extension to DFT.[74]

1.2.2.3 Force Field Approximation to the Potential Energy Surface

The calculation of stable molecular geometries, vibrational properties, dynamic simulations, or chemical kinetics parameters relies on knowledge of the potential energy surface (PES), which is derived from the electronic energy E_0 at various nuclear configurations. However, the computational cost of calculating the PES using quantum chemistry methods scales heavily with the number of atoms in the system. For instance, density functional theory (DFT) exhibits a worst-case scaling complexity of N^3 (where N is the number of basis sets for the Kohn-Sham orbitals) for E_0 calculations.[75] This makes DFT-based E_0 calculations prohibitively expensive for systems containing thousands of atoms.

To address this limitation, force field models are often employed to approximate E_0 as

the sum of covalent and non-covalent energy interactions:[75]

$$E_0 = E_{ff} = \underbrace{E_{bond} + E_{angle} + E_{torsion}}_{\text{Valence}} + \underbrace{E_{elect} + E_{vdW} + \ldots}_{\text{Non-bonded}}. \tag{1.15}$$

The exact form of Equation 1.15 depends on the specific force field used. Many force fields describe covalent energies using analytical expressions based on the harmonic approximation for bond and angle forces, and the Fourier approximation for torsional forces:[75]

$$\begin{aligned} E_{bond}(R^{AB} - R_0^{AB}) &= k^{AB}(R^{AB} - R_0^{AB})^2 \\ E_{angle}(\Theta^{ABC} - \Theta_0^{ABC}) &= k^{ABC}(\Theta^{ABC} - \Theta_0^{ABC})^2 \\ E_{torsion}(\omega) &= \sum_{n=1} V_n \cos(n\omega), \end{aligned} \tag{1.16}$$

where $(R^{AB} - R_0^{AB})$ and $(\Theta^{ABC} - \Theta_0^{ABC})$ represent the displacements from the equilibrium values for the bond $(A - B)$ and angle $(A - B - C)$, respectively; k denotes the respective force constants; ω is the dihedral angle for atoms $A - B - C - D$; V_n is the rotation barrier around the bond $B - C$; and n describes the periodicity of this rotation.[75]

The specific parameters in Equation 1.16 for bonds, angles, torsions, and non-covalent interactions must be determined through a parametrization process. This process typically relies on experimental data and/or high-level quantum chemistry calculations for systems with chemical similarity to the molecule of interest.[75,76] The parameter values in a force field depend on factors such as atomic hybridization, bonded chemical groups, electronic resonances, and other chemical properties. To account for this diversity, atom typing is often employed to classify atoms based on their bonding environments.[76]

Force field approximations enable the calculation of physicochemical parameters, structural characterization, and molecular dynamics simulations over extended timescales for systems containing thousands or even millions of atoms—tasks that are beyond the capabilities of *ab initio* methods. However, force fields based on harmonic approximations, as shown in Equation 1.16, cannot model bond-breaking processes and are therefore unsuitable for studying chemical reactions. Although specialized force fields like ReaxFF[77] can simulate chemical reactions in systems with thousands of atoms, they require a non-trivial parametrization process.

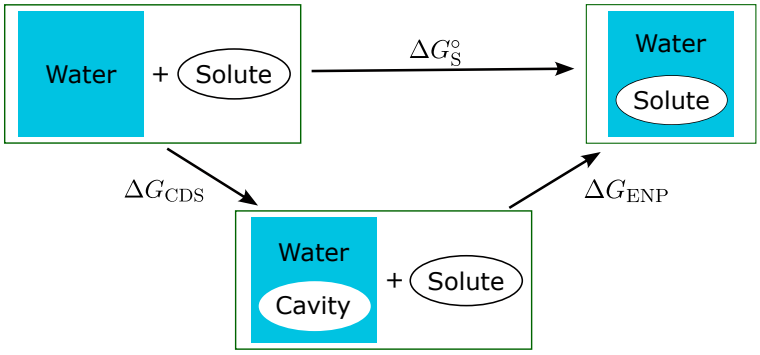


Figure 1.10: Thermodynamic cycle for the solvation process of a molecule in a dielectric continuous (structureless) solvent.

1.2.2.4 Solvent Representation

Chemical processes under physiological conditions are typically carried out in aqueous (solvent) media. This solvent environment is crucial for facilitating bond breaking and formation, and stabilizing transition states, intermediaries, and reaction products.[75,76] The equations discussed in section 1.2.2.2 describe the electronic structure of isolated molecules, akin to conditions in a vacuum or low-pressure gas phase. Therefore, it is essential to consider the interactions of reactants and products with the solvent to develop a more comprehensive computational model of chemical reactions under physiological conditions.

Incorporating a representative number of explicit solvent molecules into these calculations introduces significant computational complexity, making electronic structure calculations impractical. This challenge can be addressed by representing the solvent as a structureless dielectric continuum medium surrounding the solute molecule, which is embedded in a cavity created by the displacement of the solvent by the solute. This approach is known as the implicit solvent model.[75,76] Figure 1.10 illustrates the thermodynamic cycle for the solvation (or embedding) process of the solute in the continuous medium.

The modeling of electrostatic interaction between the solute molecule embedded in a dielectric medium relies on the Poisson equation[75,78,79]

$$\nabla\varepsilon(\mathbf{r}) \cdot \nabla\phi(\mathbf{r}) = -4\pi\varrho(\mathbf{r}). \tag{1.17}$$

In this equation, $\phi(\mathbf{r})$ represents the electrostatic potential, $\varrho(\mathbf{r})$ is the solute charge density. The electric permittivity $\varepsilon(\mathbf{r})$ is a scalar function, allowing the definition of regions in the dielectric continuum with different values of $\varepsilon(\mathbf{r})$ (see Figure 1.10). Inside this cavity, the solute

is described as fully atomistic.[79] The polarization of the dielectric medium is proportional to the electric field strength $-\nabla\phi(\mathbf{r})$.[79]

Solving Equation (1.17) requires calculating $\varrho(\mathbf{r})$, which is done using the DFT equations presented in Section 1.2.2.2. However, solute's charge distribution $\varrho(\mathbf{r})$ polarizes the dielectric continuum, and vice-versa. Given a "solute cavity" $\phi(\mathbf{r})$ is partitioned into:[79]

$$\phi(\mathbf{r}) = \phi^{\varrho}(\mathbf{r}) + \phi_{\text{rxn}}(\mathbf{r}), \tag{1.18}$$

where $\phi^{\varrho}(\mathbf{r})$ is the electrostatic potential generated by the solute's charge density, and $\phi_{\text{rxn}}(\mathbf{r})$ is a "reaction field" arising from the polarization of the continuum. In this way, the solute-solvent mutual polarization is accounted for by considering a Polarizable Continuum Medium (PCM),[79] and iteratively solving Equations (1.17) and (1.19)

$$\begin{aligned}\left[-\frac{1}{2}\nabla_1^2 - \sum_{\alpha}\frac{Z_\alpha}{r_{1\alpha}} + \int\frac{\rho(\mathbf{r}_2)}{r_{12}}d\mathbf{r}_2 + \upsilon_{ce} - \hat{\mathcal{O}}(\phi_{\text{rxn}}(\mathbf{r}))\right]\Theta^{KS}_{i,\text{solv}}(1) &= \varepsilon_i^{KS}\Theta^{KS}_{i,\text{solv}}(1),\\ \left[\hat{h}^{\text{gas}} - \hat{\mathcal{O}}(\phi_{\text{rxn}}(\mathbf{r}))\right]\Theta^{KS}_{i,\text{solv}}(1) &= \varepsilon_i^{KS}\Theta^{KS}_{i,\text{solv}}(1).\end{aligned} \tag{1.19}$$

$\hat{h}^{\text{gas}}$ is the electronic Hamiltonian of the solute in the gas-phase. The operator $\hat{\mathcal{O}}(\phi_{\text{rxn}}(\mathbf{r}))$ arises from the "reaction field" $\phi_{\text{rxn}}(\mathbf{r})$.[79] The electric permittivity $\varepsilon(\mathbf{r})$ can also be partitioned as:

$$\varepsilon(\mathbf{r}) = \begin{cases}\varepsilon_{\text{in}}, & \text{inside solute cavity,}\\ \varepsilon_{\text{out}}, & \text{outside solute cavity.}\end{cases} \tag{1.20}$$

It is usually set $\varepsilon_{\text{in}} = 1$ in quantum chemistry calculations describing the electrostatics interaction inside the cavity explicitly by the Hamiltonian $\hat{h}^{\text{gas}}$ in Equation (1.19), and ε_{out} is set according to the solvent selected (i.e, $\varepsilon_{\text{out}} = 80.4$ for water). Equation (1.19) is also a variational function for the Honenberg and Khon theorem.[75,79]

The Universal Solvent Model (SMD)[78] separates the solvation free energy ($\delta G^{\circ}_{\text{solv}}$ in Figure 1.10) into three components:

$$\delta G^{\circ}_{\text{solv}} = \Delta G_{\text{ENP}} + \Delta G_{\text{CDS}} + \Delta G^{\circ}_{\text{conc}}. \tag{1.21}$$

$\Delta G^{\circ}_{\text{conc}}$ accounts for the standard concentration change between the gas phase and the solvent media. ΔG_{ENP} accounts for the free energy associated with electrostatic (E), nuclear (N), and polarization (P) changes in the system due to the interaction between the solute and the

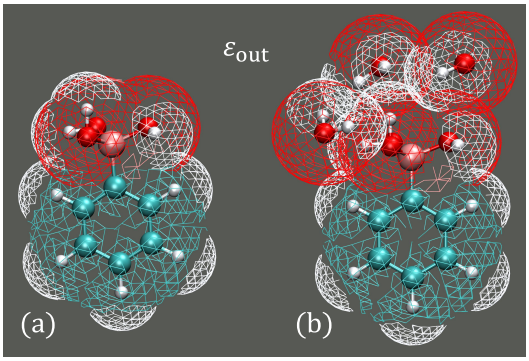


Figure 1.11: (a) PBA^{-1} molecule, and (b) a supermolecule of PBA^{-1} with three explicit water molecules in an implicit solvent (gray area) with dielectric constant ε_{out}. The mesh tessellate illustrates the solute cavity. Note the expanded solute cavity for the supermolecule in (b).

solvent.[78] If the solute geometry is practically the same in both the liquid and gas phases, the ENP term becomes an electrostatic and polarization component (EP), and ΔG_{EP} is given by

$$\Delta G_{EP} = E_v[\rho^{solv}] + \frac{e}{2}\sum_k Z_k \phi_{rxn,k} - E_v[\rho^{gas}], \tag{1.22}$$

where $E_v[\rho^{solv}]$ and $E_v[\rho^{gas}]$ are the ground state energies functionals of the electron density on the solvent media ρ^{solv} (calculated with Equations (1.17) to (1.20)) and in the gas phase ρ^{gas} (calculated with Equations (1.6) to (1.14)), respectively. Here, e is the atomic unit of charge, $\phi_{rxn,k}$ is the reaction field evaluated at atom k, Z_k is the atomic number of atom k.[78]

The term ΔG_{CDS} in Equation (1.21) is the free energy change associated with solvent cavitation (C), changes in the dispersion (D) energy, and possible changes in the solvent structure (entropy), and is given by[78]

$$\Delta G_{CDS} = \sum_k^{atoms} \sigma_k A_k(\mathbf{R}, \{R_{Z_k} + r_s\}) + \sigma^{[M]} \sum_k^{atoms} A_k(\mathbf{R}, \{R_{Z_k} + r_s\}), \tag{1.23}$$

where σ_k and $\sigma^{[M]}$ are the atomic surface tension of the atom k and the molecular surface tension, respectively. A_k is the Solvent-Accessible Surface Area (SASA) of atom k. The SASA depends on the geometry $\mathbf{R}$, the set $\{Z_k\}$ of all atomic van der Waals radii, and the solvent radius r_s. For water, $\sigma^{[M]} = 0$ and Equation (1.23) reduces to the firs term. Further details about the definitions of Equation (1.23) are found in the reference.[78]

Finally, Figure 1.11(a) exemplifies the cavity shape for a PBA molecule embedded in

a continuum solvent model. While implicit solvation models are computationally efficient for representing solute-solvent interactions, they can fail when these interactions are strong. In such cases, including some explicit solvent molecules has proven to be a cost-efficient alternative to improve the accuracy of calculated ΔG_s and related thermodynamic properties,[80–83] albeit at the expense of increased computational overhead. Figure 1.11(b) depicts a PBA molecule surrounded by three water molecules. Here, the cavity for the reaction field encompasses the interactions between the boronic groups and the explicit water molecules. Consequently, the solute interactions with the three water molecules are explicitly calculated in the Hamiltonian of Equation (1.19).

In this work, we employ various implicit solvation models, including SMD, Conductor-like Polarizable Continuum Model (C-PCM), Analitycal Linearized Poisson-Boltzmann (ALPB), and Generalized Born model augmented with the hydrophobic Solvent Accessible Surface Area (GBSA). These models differ in how they define the solute cavity's shape and size, calculate solute charge distribution and dispersion interactions, and describe the dielectric medium.[76] The selection of a specific implicit solvation model generally depends on the DFT level of theory and the availability of a compatible solvation model in the quantum chemistry software package.

1.2.2.5 Calculation of Thermodynamical Quantities from First Principles

For a reaction in solution, with a standard state "°" of 1 M, the experimental $\Delta_r G^\circ$ is calculated from the observed equilibrium reaction constant K_{obs}

$$\Delta_r G^\circ = -RT \ln(K_{\text{obs}}) = -RT \ln\left(\frac{[C][D]}{[A][B]}\right), \tag{1.24}$$

where R is the gas constant, and T is the absolute temperature of the system at equilibrium. For the chemical reaction $A + B \rightleftharpoons C + D$, the equilibrium constant can be expressed with the molar concentrations $[A]$, $[B]$, $[C]$, and $[D]$. Theoretically, the reaction-free energy can be computed according to

$$\Delta_r G^\circ_{\text{aq}} = G^\circ_{\text{aq}}(C) + G^\circ_{\text{aq}}(D) - G^\circ_{\text{aq}}(A) - G^\circ_{\text{aq}}(B), \tag{1.25}$$

from the respective absolute free energies G° of the products and reactants.

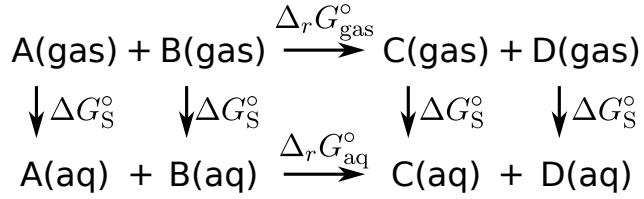


Figure 1.12: Thermodynamic cycle for the calculation of reaction free energies in aqueous phase $\Delta_r G^\circ_{\mathrm{aq}}$.

Based on the thermodynamic cycle of Figure 1.12, for a given molecule X, the free energy

$$G^\circ(X) = E_{\mathrm{gas}}(X) + G^\circ_{\mathrm{trv}}(X) + \delta G_{\mathrm{solv}}(X). \tag{1.26}$$

is calculated from the electronic energy in the gas phase $E_{\mathrm{gas}}(X)$, the free energy from the translational, rotational, and vibrational motions $G^\circ_{\mathrm{trv}}(X)$ (including the zero point energy) accessible for the molecule X at temperature T, and the solvation free energy $\delta G_{\mathrm{solv}}(X)$.

Hence, accurate computations by this brute force approach require high accuracy for all the individually computed (free) energy components. Moreover, for flexible molecules the calculation of $G^\circ(X)$ requires considering the thermodynamic average over the (at best) complete Conformer Ensemble (CE). If the flexibility of any of the reactants is likely to change due to the reaction, the conformational entropy contribution $-TS^\circ_{\mathrm{conf}}$ should be included in the molecular Gibbs free energy for the CE[84]

$$G^\circ_{\mathrm{CE}}(X) = \sum_i^{N_{\mathrm{conf}}} p_i \cdot G^\circ_i(X) - TS^\circ_{\mathrm{conf}}(X) = \overline{G}^\circ(X) - TS^\circ_{\mathrm{conf}}(X). \tag{1.27}$$

Here, $\overline{G}^\circ(X)$ is the weighted Boltzmann average over the CE where the weight for conformation i

$$p_i = \frac{e^{-G^\circ_i(X)/kT}}{\sum_j^{N_{\mathrm{conf}}} e^{-G^\circ_j(X)/kT}} \tag{1.28}$$

is calculated with the corresponding molecular free energy $G^\circ_i(X)$, the Boltzmann constant k and temperature T. j is an index that runs over the conformers (N_{conf}) used to calculate the weights p_i. By substituting eq 1.26 into eq 1.27, the molecular free energy for a CE

$$G^\circ_{\mathrm{CE}}(X) = \overline{E}_{\mathrm{gas}}(X) + \overline{G}^\circ_{\mathrm{trv}}(X) + \overline{\delta G}_{\mathrm{solv}}(X) - TS^\circ_{\mathrm{conf}}(X) \tag{1.29}$$

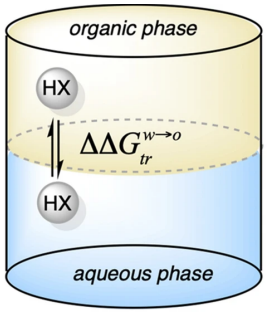


Figure 1.13: Distribution of compound HX between two immiscible solvents, octanol (o) and water (w). $\Delta\Delta G_{tr}^{w\to o}$ denotes the free energy change for the transfer of HX from water to octanol. Figure taken from reference 85.

is obtained. Finally, the reaction-free energy (Equation (1.25)) considering the CE for reactants and products is calculated by the Equation

$$\Delta_r G^\circ_{\mathrm{CE}} = \Delta\overline{G}^\circ - T\Delta S^\circ_{\mathrm{conf}}, \tag{1.30}$$

where $\Delta S^\circ_{\mathrm{conf}}$ accounts for the conformational entropy changes in the reaction.

The octanol-water partition coefficient, $\log P_{o/w}$, is a widely used physicochemical parameter to characterize the behavior of organic compounds in different solvent media. Figure 1.13 illustrates the transfer process of a compound HX from water to octanol. The partition coefficient is defined as:[85]

$$\log P_{o/w} = \frac{-\Delta\Delta G_{tr}^{w\to o}}{RT\ln 10} = -\frac{\Delta G^\circ_{\mathrm{solv},o} - \Delta G^\circ_{\mathrm{solv},w}}{RT\ln 10}, \tag{1.31}$$

where $\Delta G^\circ_{\mathrm{solv},o}$ and $\Delta G^\circ_{\mathrm{solv},w}$ are the standard solvation-free energies of the compound HX in octanol and water, respectively. R is the gas constant, and T is the temperature. If HX is a hydrophilic molecule, then $\log P_{o/w} < 0$. Conversely, $\log P_{o/w} > 0$ when HX is hydrophobic.

1.2.2.6 Chemical Accuracy

In computational chemistry, chemical accuracy refers to the accuracy needed for a model to calculate thermochemical properties for quantitative comparison with experimental values. A quantum model with chemical accuracy of $\sim$1.0 kcal/mol can reliable calculate dissociation bonds, atomization energies, and heat of formation of gas-phase molecules.[86] For chemical reactions, an accuracy of $\sim$1.0 kcal/mol in changes of free energies (ΔG°) at room temperatures ensures that calculated equilibrium and rate coefficients are within the same order of

magnitude as experimental values.

Nowadays, quantum chemical calculations can achieve and sometimes exceed the accuracy of experimental methods for thermochemical properties. Computational protocols based on wavefunction-based methods and correlation-consistent Dunning basis sets allow for a systematic approximation to the exact solution of the molecular electronic Schrödinger equation.[87] These protocols often involve computing the molecular electronic structure using the method of Coupled Cluster with Single and Double excitations and perturbative Triple excitations (CCSD(T)), which scales as $\mathcal{O}(n^3N^4)$, where n is the number of occupied orbitals and N is the number of virtual orbitals.[88] This scaling generally limits the application of composite wavefunction-based methods to calculate thermochemical properties, with accuracy < 1.0 kcal/mol, to small molecules with approximately 80-90 electrons, or medium-sized molecules like benzene.[88,89]

DFT methods scales as $\mathcal{O}(n^3)$,[89] which makes them more computationally efficient than the CCSD(T) method, allowing routine thermochemical calculations for molecules with, even, hundreds of atoms. However, numerical benchmarks have shown that the accuracy with DFT methods is limited to 2-3 kcal/mol.[88,90,91] Nevertheless, achieving chemical accuracy becomes even more challenging while increasing the molecular size because of the increasing number of electron-electron correlations, the requirement for extended basis sets to approximate the molecular electronic function,[87] and, importantly, the need for the Boltzmann averages over conformed ensembles that becomes increasingly important for flexible molecules.

To conclude this section, it is important to emphasize that the chemical accuracy of computational chemistry methods has been assessed, typically, with thermochemical experiments for relatively small molecules in the gas phase.[87,89] However, in this work, we are interested in a chemical process that took place in aqueous media. Therefore, the modeling of the solute-solvent interactions is an additional term that also influences the accuracy achieved with the selected quantum chemistry model.

In the next section, we introduce some strategies for computationally aided design and screening of molecules relevant for this work, building on the *ab initio* calculations of physicochemical properties discussed in the preceding sections.

1.2.3 *De Novo* Generation of Molecules and Evolutionary Computational Methods

The *de novo* generation of molecules is a strategy for high-throughput creation of compounds with target-specific shapes and tailored properties. Initially developed in computational drug design,[92,93] this approach has since been applied across various fields, including the development of binders for new metal-organic frameworks[94] and the creation of novel ligands for metal catalysts.[62] Figure 1.14 illustrates the general strategy employed for the *de novo* design and screening of molecules. The process begins with the generation of a large pool of candidate molecules, selected based on one or more physicochemical parameters, typically calculated using cost-effective yet reasonably accurate *in silico* methods. This initial group is then refined using *in silico* methods to calculate the properties of interest. The screening process involves several filters with increasingly strict selection criteria to identify the most promising molecules at each stage. As the refinement progresses, the accuracy and computational cost of the calculations increase. Some steps in the screening process, may involve iterative refinement of molecular properties for the best-performing candidates, using techniques such as evolutionary algorithms.[66,95] Selecting appropriate filter criteria and thresholds is critical for efficient high-throughput molecular screening. At the conclusion of the screening, a small group of molecules with optimal properties, as determined by high-accuracy methods, is selected for experimental validation.

Inverse design is a process for generating molecules based on desired properties, relying on mathematical descriptors to quantify these properties and computational algorithms to guide the design. An example of this approach was discussed in Section 1.2.1.2, where Yang and coworkers used the PBA−Glc−PBA fragment as a scaffold to design a highly selective DBA tweezer for glucose. Hay and coworkers[60,61] developed the HostDesigner suite of programs for inverse design and *de novo* generation of organic host molecules tailored to the chemical and geometrical requirements of metal ions or small organic molecules. The host molecules are created using a database of molecular fragments and codified chemical bonding rules.[60,61] While the generated ligands meet the geometric requirements of the guest molecule, their geometry may not necessarily result in a stable conformation within the host cavity. This stability is verified through a conformational search of the generated ligands.

More recently, *de novo* generative methods has been developed through machine learning algorithms.[93,96,97] However, this requires the prior knowledge of a significant amount of high quality data, from experiments or computational calculations, of closely related systems

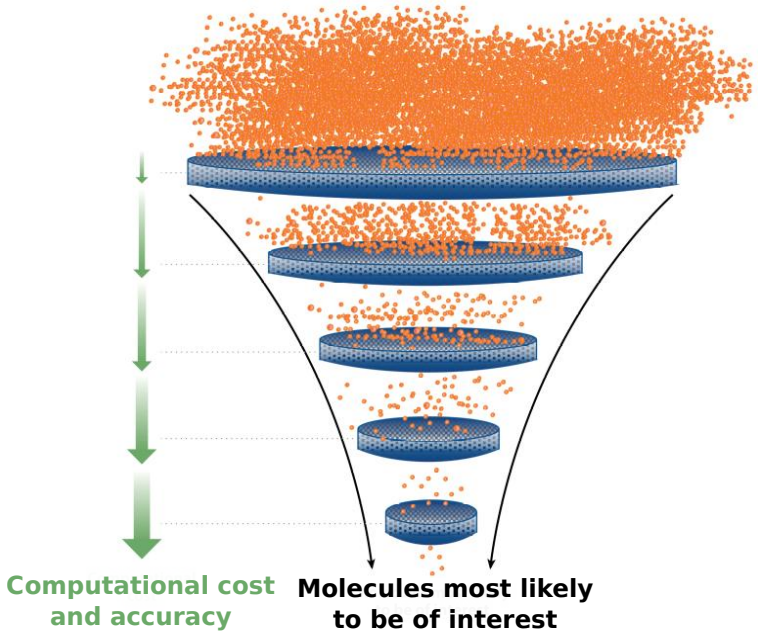


Figure 1.14: A computational funnel approach: progressively stricter filtering criteria are applied to exclude uninteresting molecules, narrowing down the virtual library to the top-performing candidates. Figure taken from reference 92.

to train and test deep neuronal networks,[96,97] which may limit the application of machine learning methods to some cases. In this regard, evolutionary computing methods are an appealing alternative for *de novo* and inverse design of molecules because they do not require prior information of the problem, and some authors have reported evolutionary computational methods can surpass the efficiency and general performance of machine learning algorithms for *de novo* generation of molecules.[66,98,99]

The evolutionary computing methods use the Darwinian natural evolution selection rules to find the best approximation to the global solution of complex problems.[100] These methods use a population of possible solutions, from which the best individuals (or best solutions) are modified iteratively through evolutive operations such as reproduction, mutation, and recombination until the iteration cycles are exhausted or a convergence criterion is achieved.[100] Genetic programming is a branch of evolutionary computing where the population of individuals consists of computer programs that solve a particular task. For example, a program might consist of a sequence of arithmetic operations that together form a mathematical expression to solve an algebraic equation.[101] The fittest program solution is the one that better solves the problem. Once a code programming language is selected, a population of computer programs is modified iteratively through genetic operations creating new programs.[101]

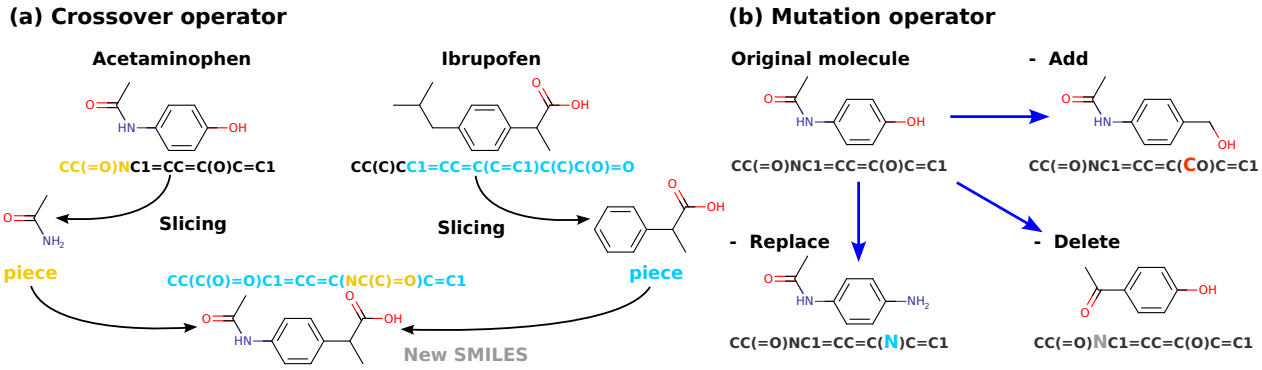


Figure 1.15: Genetic operations (a) crossover and (b) mutation for *de novo* generation of molecules based on genetic programming. The text below the molecules corresponds to the Simplified Molecular Input Line Entry System (SMILES) string to codify the graph molecule. Figure taken from reference 102.

Genetic programming has been applied to the *de novo* design of molecules using the Simplified Molecular Input Line Entry System (SMILES).[66,95,102–104] SMILES is a chemical notation that represents molecular graphs of small molecules as strings of ASCII characters, capturing essential features such as atom chirality, charges, bond types and orders, rings, and branches.[105] Its string-based format enables efficient machine processing and manipulation of molecular structures, establishing SMILES as a standard in cheminformatics. For example, SMILES is widely adopted by cheminformatics libraries like RDKit,[106] which includes modules for modifying SMILES strings, performing substructure searches using the SMILES arbitrary target specification (SMARTS),[107] and making modifications based on notations similar to standard chemical reactions.[106]

Figure 1.15 illustrates an evolutionary algorithm that applies crossover and mutation operations on SMILES strings to generate new molecular graphs. In the crossover operation (Figure 1.15 (A)), SMILES strings from two parent molecules are cut and recombined to form a new offspring molecule. In the mutation operation (Figure 1.15 (B)), the SMILES string is altered by adding, replacing, or deleting atoms (nodes) or bonds (edges), thereby generating a new molecule.[102] To maintain a well-defined optimization problem, it is essential to limit the molecular size of the newly generated molecules.

Chapter 2

Problem Statement and Objectives

2.1 Problem Statement

A deep understanding of sucrose's role as a plant biomarker is crucial for breeding plant varieties with enhanced genetic and phenotypic traits to better withstand biotic and abiotic stresses, which are increasingly frequent and severe due to global warming. Currently, the lack of technologies capable of monitoring sucrose concentrations in real crop-field conditions impedes the identification of specific genes involved in the plant's stress responses regulated by sucrose.

A technology with the capability to quantify sucrose in plants, under real conditions, should meet several requirements such as i) high selectivity, ii) no destructive measurement, iii) no disturbance of the plant's homeostatic conditions, iv) portability for operation in crop fields, v) high spatiotemporal resolution. Nowadays, the detection and quantification method of sucrose is one of the main obstacle to developing a method for quantification of sucrose with the abovementioned requirements. Most biosensors react irreversibly with sucrose, operate under high specific reaction conditions, and potentially disturb the plant equilibrium. On the other hand, the analytical chemistry methods require sample preparation, robust, expensive, and complex facilities for machine operation, incompatible with the required high spatiotemporal resolution and crop-field measurements.

Tweezer-like DBA molecules have demonstrated some success in the selective quantification of glucose under *in vivo* conditions. However, there is no general procedure to extend the advantages of these molecules to the quantification of other carbohydrates. The *in silico* inverse design methodology proposed by Yang and coworkers,[17] while (potentially) capable of designing selective tweezer DBA molecules for specific carbohydrates, has not been applied

to create novel DBA architectures for carbohydrates other than glucose. This limitation may arise from the structure-based nature of Yang's approach, which does not adequately account for factors such as repulsive electrostatic interactions, high-energy conformations, or unfavorable entropy changes that could hinder spontaneous DBA-carbohydrate binding despite apparent geometric complementarity. Additionally, Yang's approach does not account for synthetic accessibility during the tweezer design process, potentially leading to complex and time-consuming synthesis routes that result in impractical tweezer designs for experimental implementation. Consequently, identifying a selective DBA for a specific carbohydrate often remains a matter of serendipity.[18]

2.2 Objectives

In this dissertation, we address the challenge of quantifying sucrose under physiological conditions by designing novel tweezer-like DBAs molecules with improved reactivity and selectivity toward sucrose, by careful and precise positioning of the boronic groups for reactions with the hydroxyl groups of sucrose. The molecular structures of DBA should be relatively rigid, yet with sufficient rotational freedom to adjust the mutual orientation of both boronic groups and facilitate their interactions with sucrose.

2.2.1 General Objective

In silico design, characterize, and screen low molecular weight ligands with high selectivity towards sucrose, and validate best performers using first-principles quantum mechanics methods.

2.2.2 Specific Objectives

1. Study existing methods that have been shown selective to sucrose detection and quantification.

2. *In silico* design, characterize and screen molecular-scale ligands with high affinity to sucrose, based on most promising sensing method identified from objective 1.

3. Characterize the physicochemical changes generated by the tweezer-sucrose binding, which can be interpreted as a signal for the quantification of sucrose (e.g. conformational, vibrational, optical, or other changes).

4. Propose a synthesize route, in collaboration with an organic synthesis group, of at least one candidate from the group of best-performing candidates obtained from the in-silico screening process.

5. Suggest a high sensitivity and selectivity detection and quantification method using the best performing sucrose ligand obtained from *in silico* screening.

Chapter 3

Methodology

In this work, we proposed a holistic approach to designing molecular tweezers for selective sucrose detection. A state-of-the-art computational chemistry workflow[108] allowed us to calculate the reaction-free energies considering the conformational entropy changes (see section 1.2.2.5) and the strong solute-solvent interactions (see section 1.2.2.4) for the reaction of tweezer-like DBA molecules with carbohydrates. By exploring the chemical space using evolutionary computational techniques within a multi-parametric optimization framework, we modeled synthetically accessible DBAs with favorable physicochemical properties for the detection of sucrose in aqueous media. The following sections outline our computational workflow.

3.1 Calculation of Free Energies for Reaction of Diboronic Acid Molecules with Monosacharides

Carbohydrates are highly flexible and soluble molecules in water.[109] Meanwhile, tweezer-like DBA molecules exhibit the highest reactivity with carbohydrates when the boron atom carries a formal negative charge and adopts sp^3 hybridization.[19,45] This makes modeling solute-solvent interactions between DBAs and water particularly challenging for implicit solvation models (section 1.2.2.4). These factors complicated the accurate calculation of reaction-free energies, as thermodynamic properties for microsolvated reactants and products must be averaged over a CE, significantly increasing computational costs. In this section, we present our computational workflow for calculating reaction-free energies while managing these costs.[108] The accuracy of this workflow is validated by comparing the calculated $\Delta_r G^\circ$

of DBA molecules with glucose and galactose to experimental values.

3.1.1 Conformational Sampling and Molecular Free Energies of Microsolvated Molecules

The conformational space of reactants and products for the reactions are sampled with the Conformer-Rotamer Ensemble Sampling Tool (CREST)[110] (version 2.11.2) while using the included Quantum Cluster Growing (QCG)[111] method for microsolvation with water. Analysis of Autocorrelation Functions (ACFs) (see Figures A.2 to A.5, A.8 to A.11, A.13, and A.14) were used to set the length of Metadynamics (MTD) simulatins to 80 ps for all calculations, to ensure a well-converged sampling of the conformational space for these complicated systems. Subsequently, a refined energetic sorting of the obtained ensemble is calculated with the Commandline Energetic SOrting (CENSO) tool[84] (version 1.2.0). Here, the refinement can be divided into parts $0-3$, where all parts allow the individual setup of the used Density-Functional Approximation (DFA) for the electronic energy and the threshold for sorting out the high-lying conformers. In part 0, a cheap prescreening is performed based on the electronic energy from a single-point (SP) calculation. Further prescreening in part 1 uses free energies by including thermostatistical contributions[112] with the GFN2-xTB[ALPB] method, which is an extended semiempirical tight binding approach that includes anisotropic second-order density fluctuation effects via short-range damped interactions of cumulative atomic multipole moments and an ALPB model to account for solvation effects.[113] In part 2, the remaining structures are optimized at the chosen level of theory. Finally, part 3 performs single-point calculations on the previously optimized structures. The choice of functional and sorting thresholds for the different parts will be discussed in the results section. While implicit solvation is included in part 0 via the GBSA model,[114–116] the succeeding parts use the SMD solvation model.[78] Both, explicit and implicit water models are used in the conformer ensemble generation with QCG.

3.1.2 Calculation of Reaction-Free Energies with Conformer Ensembles

Since the calculation of molecular free energies for flexible systems based on meta-dynamics involves randomness, multiple trials (N) are needed to adequately sample the value distribution of $G^{\circ}_{\mathrm{CE}}(X)$. With $G^{\circ}_{\mathrm{CE,i}}(X)$ being one of the sampled molecular free energies, for a par-

ticular permutation of the group $\{i, j, k, l\}$ (corresponding to the index $x \in \{1, 2, 3, ..., N^4\}$) among the N-sampled molecular free energies for the reactants and products, the Equation 1.25 turns into

$$\Delta_r G^\circ_{\mathrm{CE},x} = G^\circ_{\mathrm{CE},k}(C) + G^\circ_{\mathrm{CE},l}(D) - G^\circ_{\mathrm{CE},i}(A) - G^\circ_{\mathrm{CE},j}(B), \tag{3.1}$$

with $i, j, k, l \in \{1, 2, 3, ..., N\}$. As the choice of $\{i, j, k, l\}$ indices is arbitrary, we defined the reaction-free energy

$$\Delta_r G^\circ_{\mathrm{CE}} = \mathrm{med}\left(\{\Delta_r G^\circ_{\mathrm{CE},x}\}\right). \tag{3.2}$$

as the median of the distribution of the set of values $\Delta_r G^\circ_{\mathrm{CE},x}$, ordered from smallest to greatest and calculated for all possible combinations of $\{i, j, k, l\}$ indices. Now, the precision of the workflow can be defined as the IQR of this distribution.

The thermostatistical contributions $G^\circ_{\mathrm{trv}}(X)$ are calculated with semiempirical GFNn-xTB methods, and the implicit solvation method ALPB[113] is used for calculation of solvation free energy $\delta G_{\mathrm{solv}}(X)$. While $\overline{E}_{\mathrm{gas}}$, $\overline{G}^\circ_{\mathrm{trv}}$, and $\overline{\delta G}_{\mathrm{solv}}$ from Equation 1.29 are automatically calculated by CENSO with the given settings. The conformational entropy S°_{conf} is determined by the ensemble entropy calculation from CREST based on GFN2-xTB generated ensembles with the GBSA implicit solvent model for water.

3.1.3 Experimental Reference Data

To assess the accuracy of our workflow for calculation of reaction-free energies, we selected the reactions **DBA-1**$^{-2}$+Glc (Figure 3.1), **DBA-2**+Glc (Figure 3.2), and **DBA-1**$^{-2}$+Gal (Figure 3.3) because the structure of reactants and products has been correctly characterized. Furthermore, the structures of the products in the reactions **DBA-1**$^{-2}$+Glc and **DBA-2**+Glc have been elucidated by analyzing the Nuclear Magnetic Resonance (NMR) spectra of proton (^{1}H) and carbon (^{13}C) nuclei, and by determining dipolar couplings ($J_{\text{C-C}}$ and $J_{\text{H-H}}$) to accurately identify the hydroxyl groups of glucose bonded to the boron atoms in the DBAs.[55,117]

The reactions **DBA-1**$^{-2}$+Glc, **DBA-2**+Glc, and **DBA-1**$^{-2}$+Gal were performed at conditions where the solution's pH > pK_as of the DBAs (see Table 3.1). In the Appendix A, we show that the equilibrium side reactions for reactants and products are unimportant at the corresponding reaction's pH. Then, the equilibrium reaction constants K_{obs} and the experimental reaction-free energies $\Delta_r G^\circ_{\mathrm{Exp}}$ can be safely assigned to the chemical processes shown

Table 3.1: Acidity constants (pK_a) and the corresponding solution's pH reported for the calculation of the observed equilibrium constants K_{obs} (L mol^{-1}) for the reactions of **DBA-1**$^{-2}$+Glc, **DBA-2**+Glc, and **DBA-1**$^{-2}$+Gal. α is the speciation fraction of the corresponding reactant at the given reaction pH's. The experimental reaction-free energies $\Delta_r G^\circ_{Exp}$ (kcal/mol) were calculated with eq 1.24.

	pK_{a1}	pK_{a2}
DBA-1^{-2}	8.5[a]	9.5[a]
DBA-2	3.7[55]	4.7[55]
Glc[118,119]	12.1	13.9
Gal	∼12.0	–

Reaction	pH	α^a_{Sugar}	α^a_{DBA}	K_{obs}	$\Delta_r G^\circ_{Exp}$
DBA-1^{-1}+Glc	11.3	0.89	0.98	19 000[b]	−5.8
DBA-2+Glc	7.4	1.0	1.0	2510±1.2[c]	−4.6 ± 0.1
				10 000[d]	−5.5
DBA-1^{-1}+Gal	11.3	∼0.90	0.98	2200[b]	−4.2

[a] See the Supporting Information for the calculation of these values.
[b] K_{obs} calculated with circular dichroism spectroscopy using the Benesi-Hildebrand method, and assuming a reaction yield of 100% for the products **DBA-1.Glc**$^{-2}$ and **DBA-1.Gal**$^{-2}$ in references 120 and 53.
[c] K_{obs} from potentiometric and fluorescence titrations.[55]
[d] K_{obs} from NMR spectroscopy.[55]

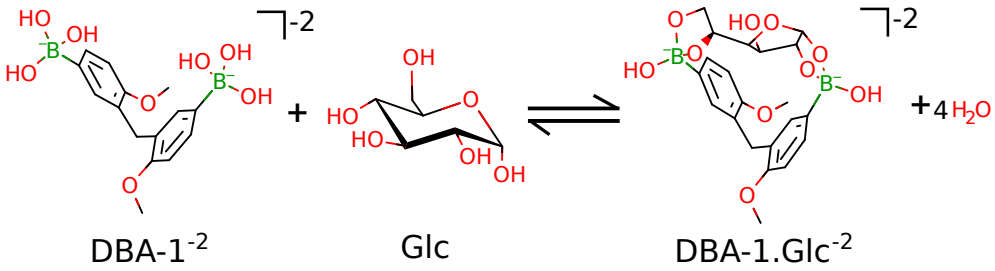


Figure 3.1: Reaction of **DBA-1** with glucose in aqueous alkaline solution at pH = 11.3.[117]

in Figures 3.1, 3.2, and 3.3. Consequently, the reference reaction-free energies $\Delta_r G^\circ_{Exp}$ are calculated directly from the corresponding experimental K_{obs} using Equation 1.24 and are presented in Table 3.1. Hereafter, we will also use the symbols $\Delta_r G^\circ_{CD}$, $\Delta_r G^\circ_{PT}$, and $\Delta_r G^\circ_{NMR}$ to refer to the experimental reaction-free energies ($\Delta_r G^\circ_{Exp}$) calculated from the equilibrium constants (K_{obs}) from Circular Dichroism (CD), Potentiometric Titration (PT), and NMR experiments, respectively.

In all our calculations we used the α-glucopyranose and α-galactopyranose isomers because the 1,2-hydroxyls in cis position (see Figures 3.1, 3.2, and 3.3) make these isomers the most reactive towards **DBA-1**$^{-2}$ and **DBA-2**. Meanwhile, in the β-isomer, the 1,2-diols are in trans positions which do not form a stable five-member diboron ester ring.[45]

Despite the reported value[53,120] of the equilibrium constant K_{obs} for the reaction of **DBA-1**$^{-2}$+Gal (Figure 3.3), the structure of the product **DBA-1.Gal**$^{-2}$ has not yet been charac-

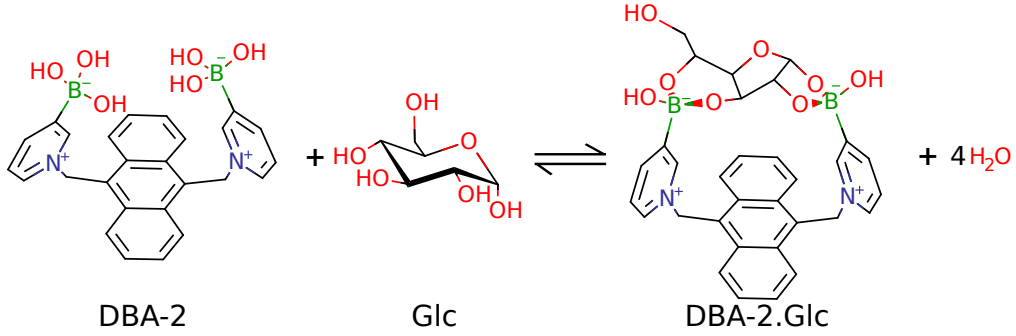


Figure 3.2: Reaction of **DBA-2** with glucose in aqueous solution at pH = 7.4.[55]

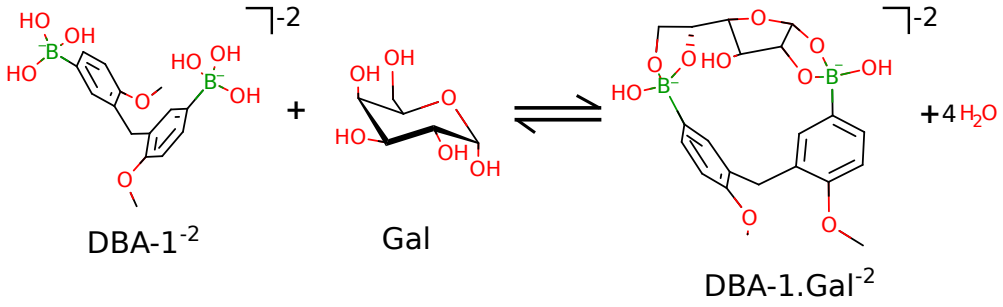


Figure 3.3: Proposed structures for the reaction of **DBA-1**$^{-2}$ with galactose at the conditions reported in reference in 117 (aqueous solution at pH = 11.3).

terized. Nevertheless, the structural and chemical similarities between glucose and galactose allow us to propose a reasonable structure for **DBA-1.Gal**$^{-2}$, as shown in Figure 3.3, for our calculations of the reaction-free energy. Additionally, the reaction conditions for **DBA-1**$^{-2}$+Gal (pH = 11.3) allow us to discard the side equilibrium reaction and use similar speciation fractions α as in the case of **DBA-1**$^{-2}$+Glc reaction, as shown in Table 3.1.

Finally, the K_{obs} were calculated with diluted solutions. For the **DBA-1**$^{-2}$+Glc and **DBA-1**$^{-2}$+Gal reactions,[53,120] the initial molar fractions of the reactants range from $4.5\text{x}10^{-5}$ to $7.2\text{x}10^{-5}$. In the case of **DBA-2**+Glc reaction,[55] the initial molar fractions range from $4.5\text{x}10^{-7}$ to $4.5\text{x}10^{-3}$. At equilibrium, these molar fractions decrease further due to the production of four water molecules in the reaction. Thus, the systems studied in this work (Figures 3.1, 3.2, and 3.3) can be closely represented by our computational workflow. Therefore, solute-solute interactions are expected to have minimal weight in the discrepancies between our calculated reaction-free energies ($\Delta_r G^{\circ}_{\text{CE}}$) and experimental results ($\Delta_r G^{\circ}_{\text{Exp}}$).

3.1.4 Systematic Screening of Computational Parameters

The technical parameters of the methods used in this work (QCG, CREST, and CENSO) form a vast space, hence identifying the optimal combination for calculating the reaction-free energies $\Delta_r G^\circ_{\mathrm{CE}}$ is a complex endeavor. Additionally, the determination of the number of water molecules required to refine the representation of solute-solvent interactions poses another parameter to optimize. Nevertheless, the computational cost increases and the conformational sampling becomes more complicated with the addition of explicit water molecules for microsolvation. Hence, we performed a parameter optimization step for the calculation of the reaction-free energies $\Delta_r G^\circ_{\mathrm{CE}}$ for the reactions of **DBA-1**$^{-2}$+Glc (Figure 3.1), **DBA-2**+Glc (Figure 3.2), and **DBA-1**$^{-2}$+Gal (Figure 3.3).

The parameter search comprised two phases. The first phase involved optimizing the CENSO parameters to compute the energy thresholds needed to sort out energetically unfavorable conformers. Table 3.2 tabulates the combinations of theory levels and the corresponding energy thresholds (G_i^{thr}) used in the first phase of the parameter search. Each combination was selected taking into account that the computational cost increases when moving forward in the CE refinement with CENSO. That excludes settings D where we were interested in the performance of the HF-3c method for optimization. For the first phase of the parameter search, we only use the reaction of the smaller **DBA-1**$^{-2}$ with glucose (Figure 3.1) to reduce computational costs during the parameter search.

The 3c methods were chosen due to their good cost-performance ratio, needed for screening a large amount of highly flexible systems[121–124] while using a non-composite functional like B97-D3 allows increasing the basis set size with the CENSO parts. The B97-D3 functional is used with the Atomic Orbital (AO) basis set def2-SV(P) in part 0, and the def2-TZVP(-f) AO basis in parts 1 and 2. Concerning this, the 3c methods have been designed to accurately calculate Non-Covalent Inteaction (NCI) using a tailor-made basis set. The accurate description of non-covalent interactions is especially important when microsolvation is considered. For these reasons, different combinations of the methods HF-3c, B97-D3, and r^{2}SCAN-3c are evaluated.

The second phase of the parameter search screens the number of water molecules necessary to achieve accurate free energies $\Delta_r G^\circ_{\mathrm{CE}}$ for the reactions of **DBA-1**$^{-2}$ and **DBA-2** with glucose, using the best-performing combination of DFA methods from the first phase. The number of water molecules for microsolvation was gradually increased in all the reactants and products, to assess the precision, accuracy, and robustness of the DFA methods in CENSO for calculating $\Delta_r G^\circ_{\mathrm{CE}}$. Although most of the QCG and CREST parameters were left as

Table 3.2: Settings for CENSO calculations with assigned letters for referencing. Columns part 0 - 3 show the functional used in the cheap prescreening, prescreening, optimization, and refinement parts of the program. The energy windows used for screening out are given by G_i^{thr} with the index corresponding to the different parts.

Set.	part 0	part 1	part 2	part 3	G_0^{thr} $[\frac{\text{kcal}}{\text{mol}}]$	G_1^{thr} $[\frac{\text{kcal}}{\text{mol}}]$	G_2^{thr} $[\frac{\text{kcal}}{\text{mol}}]$
A	B97-D3	r^{2}SCAN-3c	r^{2}SCAN-3c	-	6	4	2.5
B	B97-D3	r^{2}SCAN-3c	r^{2}SCAN-3c	-	10	6	3
C	B97-D3	r^{2}SCAN-3c	r^{2}SCAN-3c	-	15	10	6
D	HF-3c	HF-3c	HF-3c	-	15	10	6
E	HF-3c	B97-D3	B97-D3	-	15	10	6
F	B97-D3	HF-3c	HF-3c	r^{2}SCAN-3c	15	10	6
G	B97-D3	B97-D3	B97-D3	-	15	10	6

default, all the meta-dynamics simulations for conformational sampling were conducted for 80 ps. To avoid any bias in the conformational sampling, a new CE was generated for all the reactants and products for each calculation of the reaction-free energy.

In the next sections, we present the parts of our workflow related whit the computational design of tweezer-like DBA molecules for selective reaction with sucrose.

3.2 Screening Stable Products from the Reaction of Sucrose with aBA Molecules

The selectivity of aBA-based molecules towards a target carbohydrate depends on the distance and orientation of the boronic groups to react, preferentially, with the hydroxyl groups in the carbohydrate.[45] However, no experimental reports exist about the simultaneous reaction of sucrose with two aBA molecules forming four boron-diester bonds. For this reason, we performed *ab initio* calculations to screen the most stable products for the simultaneous reaction of two aBA molecules with sucrose in aqueous media. In these structures, the aBA molecules are linked to sucrose by forming boronate diester bonds with hydroxyl groups at either vicinal 1,2- or 1,3-positions (see Figure 1.1). Figure 3.4 show the generic scheme for sucrose reacting with two PBA molecules.

The most stable product was assessed from the reaction Enthalpy ($\Delta_r H^\circ$) in aqueous media for the equilibrium reaction in Figure 3.4. $\Delta_r H^\circ$ is given by the equation:

$$\Delta_r H^\circ = H^\circ_{aq}(\text{aBA}^{-1}\text{-Suc-aBA}^{-1}) + 4H^\circ_{aq}(\text{H}_2\text{O}) - H^\circ_{aq}(\text{Suc}) - 2H^\circ_{aq}(\text{aBA}^{-1}). \tag{3.3}$$

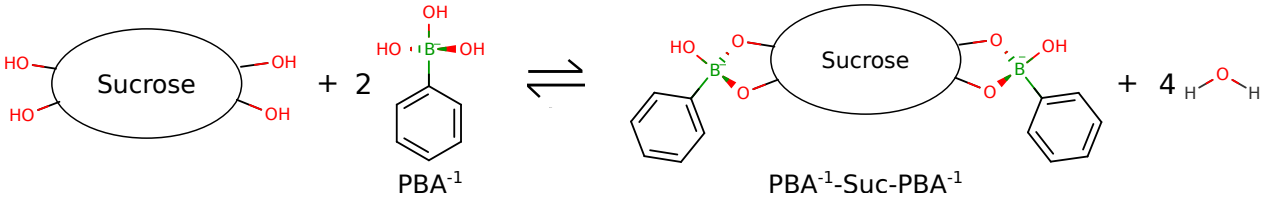


Figure 3.4: Generic reaction scheme for screening the most reactive hydroxyl (-OH) groups in sucrose for the formation of boron-diester bonds with PBA.

For any of the reactant and products in Figure 3.4, the molecular enthalpy in aqueous media is $H^{\circ}(X)$ is given by the equation:

$$H^{\circ}(X) = E_{\text{gas}}(X) + \delta G_{\text{solv}}(X). \tag{3.4}$$

Geometry optimizations for both reactants and products were carried out using the composite DFT method PBEh-3c,[125] which expands the molecular orbitals in an Ahlrichs-type valence-double zeta AO Gaussian basis set, incorporates geometric counterpoise correction to account for basis set superposition error, and includes two- and three-body dispersion corrections.[125] The aqueous solvent environment was modeled using the SMD implicit solvation model.[78] All the calculations for this section were performed with the ORCA software suite.[126]

3.3 Reaction Mechanism of aBA with Sucrose

To determine the Minimum Energy Path (MEP) for the reaction mechanism of aBAs with sucrose, the Nudged Elastic Band-Climbing Image (NEB-CI) method[127,128] was employed to explore the PES. To simplify the problem, we examined theMEP for the reaction of one aBA molecule with the hydroxyl groups in the glucose and fructose units of sucrose, excluding the binding of a second aBA molecule to sucrose. The reactant was modeled as the supermolecule aBA + Suc, while the product was the borondiester aBA−Suc complex with water molecules, aBA−Suc.$(H_2O)_2$, as illustrated in Figure 3.5. The relative positions of the aBA and sucrose in the reactants, as well as the water molecules in the product, were optimized to ensure a smooth interpolation of sixteen configuration images for the initial reactive path between the reactant and product. The image interpolation was performed with the image-dependent pair potential[129] implemented in ORCA.[126]

The highest-energy image from the converged NEB-CI calculations was used as the starting geometry for Transition State (TS) optimization to locate the first-order saddle point

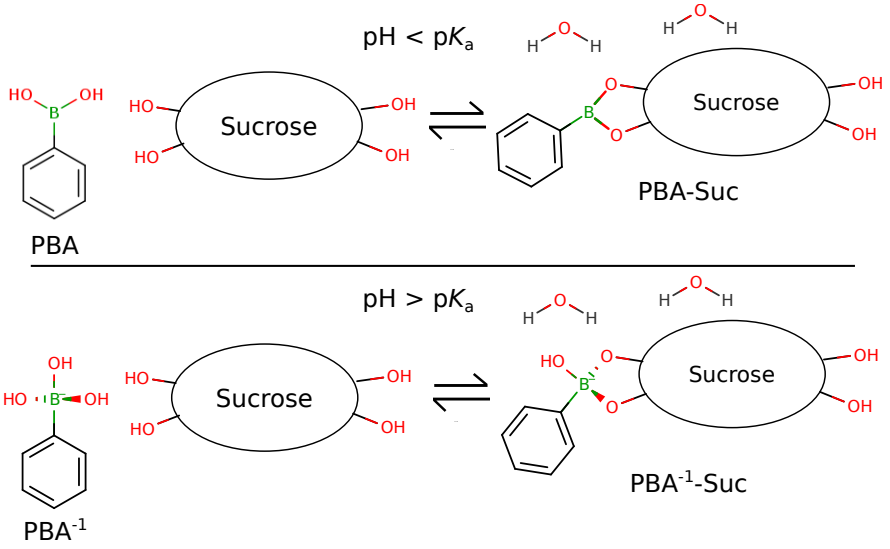


Figure 3.5: Reaction scheme of PBA with sucrose under different pH conditions, explored using NEB-CI for the MEP, followed by Transition State (TS) optimization, frequency analysis, and Intrinsic Reaction Coordinate (IRC) calculations. The pK_a represents the acidity constant of PBA molecule.

along the MEP for the corresponding reactions. A frequency calculation was then performed on the optimized transition state to verify the presence of a single negative frequency in the Hessian matrix, confirming the first-order saddle point. Finally, Intrinsic Reaction Coordinate (IRC) calculations were conducted to ensure the optimized transition state correctly connects the reactants and products.

The literature suggests that aBAs are more reactive toward carbohydrates under conditions where the pH $>$ p$K_{a,\mathrm{aBA}}$.[45] In this scenario, the dominant aBA specie corresponds to the boronate form R$-$B$(\mathrm{OH})_3^{-1}$ (also denoted as aBA^{-1}). However, aBA can also react with carbohydrates under suboptimal reaction conditions, where the pH $<$ p$K_{a,\mathrm{aBA}}$, and the neutral form of aBA (R$-$B$(\mathrm{OH})_2$) becomes the dominant species. We investigated the MEP for reaction of aBA with sucrose under both reaction conditions pH $>$ p$K_{a,\mathrm{aBA}}$ and pH $>$ p$K_{a,\mathrm{aBA}}$, as shown schematically Figure 3.5.

The NEB-CI and geometry optimizations in this section were carried out using the DFA method r^{2}SCAN-3c,[130] with the SMD implicit solvent model for water. Given that r^{2}SCAN-3c is susceptible to self-interaction errors,[90] reaction barriers were recalculated with single point calculations at the ωb97M-D4/def2-QZVPP level of theory, using molecular geometries of the critical points on the PES obtained from r^{2}SCAN-3c. All calculations were performed using the ORCA[126] software suite.

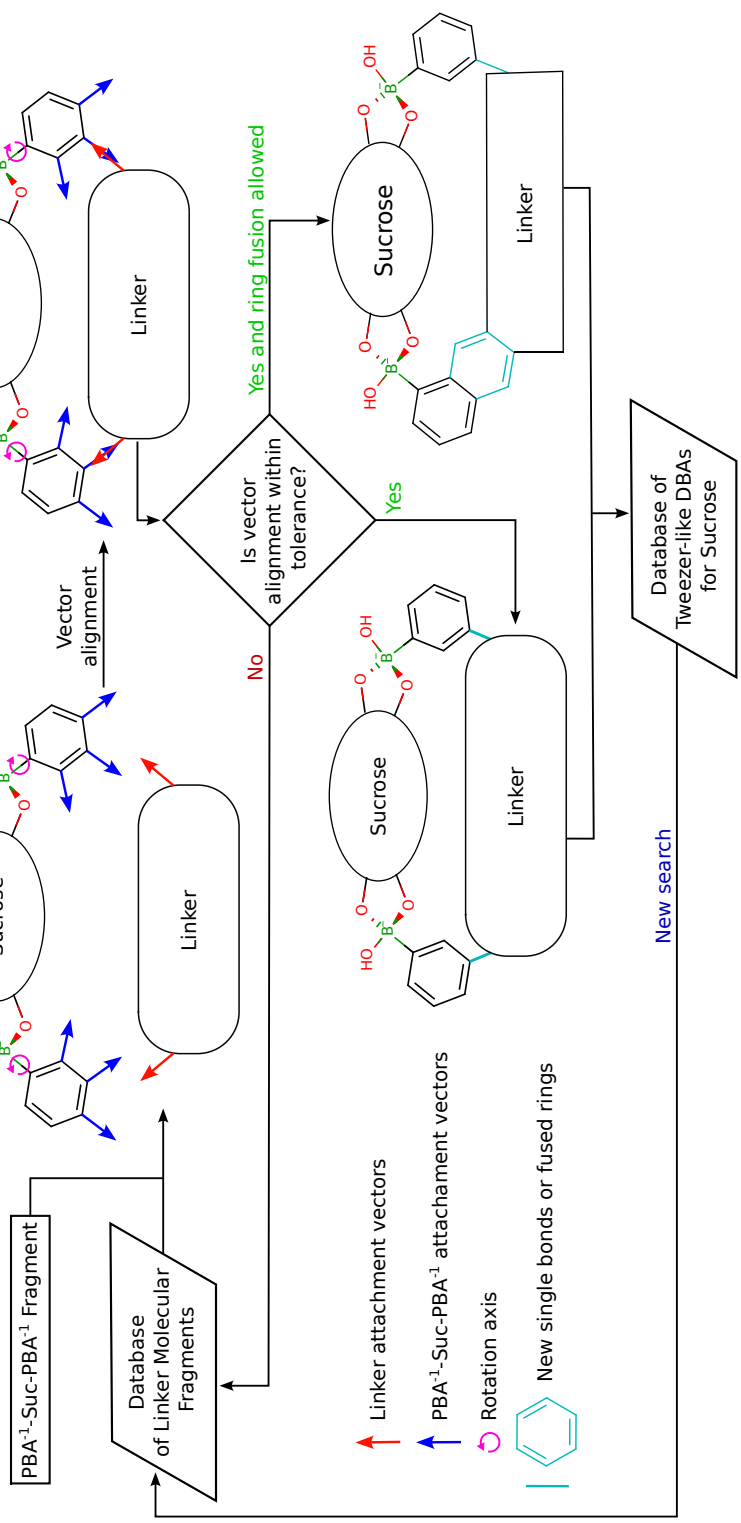


Figure 3.6: Workflow for the *de novo* generation of tweezer-like DBA molecules targeting sucrose, using the OVERLAY program[60] from the HostDesigner suite.[61] The tails of the attachment vectors correspond to carbon atoms, while the heads correspond to hydrogen atoms.

3.4 *De Novo* Generation of Tweezer-Like DBA Molecules for Sucrose

The most stable $aBA^{-1}-Suc-aBA^{-1}$ products (as shown in Figure 3.4) were used as scaffolds for the inverse design of new tweezer-like DBA molecules. In the $aBA^{-1}-Suc-aBA^{-1}$ scaffold, the aBA moieties already have the optimal distance and orientation for reaction with the hydroxyl groups of sucrose. The aBA moieties were locked in this ideal position by a molecular fragment that closed the tweezer structure. The OVERLAY program[60] from the HostDesigner[61] suite was used to find the best matching molecular fragment to close the tweezer structure.

The OVERLAY program requires the definition of attachment vectors where the linker binds to close the tweezer. These vectors (blue arrows in Figure 3.6) were aligned colinearly with the C−H bonds at the *ortho-*, *meta-*, and *para-* positions relative to the boronic group in the $aBA^{-1}-Suc-aBA^{-1}$ scaffolds. OVERLAY finds a matching molecular fragment based on the alignment between the scaffold's attachment vectors and linker's attachment vectors (red arrows in Figure 3.6). This alignment was assessed by comparing the distances between the tail vectors of the scaffold and the head vectors of the linker, and vice versa, as well as by comparing the dihedral angle of the attachment vectors in the scaffold with those in the linker fragment. Smaller differences in these distances and dihedral angles indicate better complementarity between the molecular fragments. $aBA^{-1}-Suc-aBA^{-1}$+linker fragments with a Root Mean Square Displacement (RMSD) ≤ 3.0 Å for vectors distance, and dihedral angles differences $\leq 60.0°$, were covalently bonded and saved in the database of tweezer-like DBA molecules for sucrose. A key feature of HostDesigner (version 4.1) is its ability to merge the $aBA^{-1}-Suc-aBA^{-1}$ scaffold with the linker either through single bonds or by atom and bond fusion within the linker's rings, when applicable, as illustrated in Figure 3.6. Further details about the fusion rules can be found in the HostDesigner program manual (version 4.1).

HostDesigner's DRIVE algorithm generates new template structures by adjusting the link angles and dihedrals of the original template. Rotations around the B−C bonds to adjust the phenyl rings and around the glycosidic bonds in sucrose were performed to generate more tweezer structures and improve their quality. The rotation axes for the phenyl rings are depicted by magenta curly arrows in Figure 3.6. The phenyl rings were rotated in 40.0° increments, from 0.0° to 350.0°. Smaller increments were tested, but they did not enhance the results and significantly increased the computational cost of generating the tweezer-like

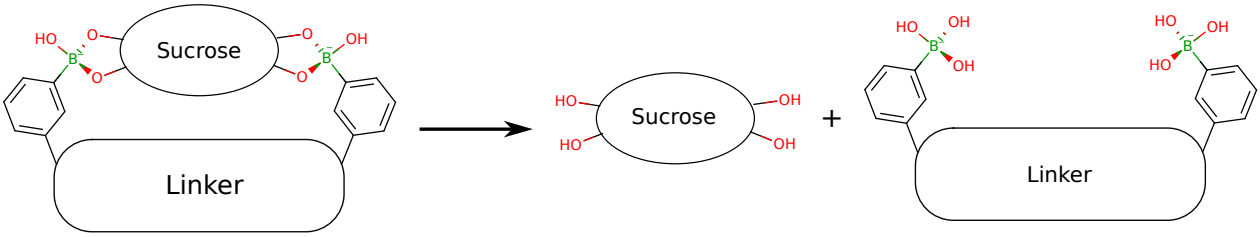


Figure 3.7: Separation of the Sucrose-DBA complex molecule for conformational ranking of the molecules generated with HostDesigner.

structures. Rotations around the glycosidic bond of sucrose were applied to the glycosyl side in 5.0° increments, ranging from -35.0° to 5.0°. For the fructosyl ring, rotations were made in 1.0° increments, from -10.0° to 10.0°. All rotations were performed relative to the input structure.

The linker fragment database from the HostDesigner software suite (version 4.1)[61] was used to identify the best matching fragments to close the tweezer-like molecules. These fragments were subjected to molecular optimization and conformational searches using the MM3 force field to identify the most stable conformers. Additional details about the design of the linker database can be found in references 60 and 61.

All the aforementioned operations were performed iteratively using fragments from the linker fragment database of HostDesigner to close the tweezer. After completing these iterations, a set of sucrose molecules that bonded to the tweezer DBA was generated and stored in a database, all within the specified RMSD and dihedral tolerances.

The database of sucrose-DBA complexes was refined through geometry optimizations using the PBEh-3c level of theory[125] and the C-PCM model[131] for implicit representation of the aqueous environment. The refinement process involved optimizing the geometry of the sucrose-DBA complex, separation into its constituent molecules (sucrose and the DBA tweezer), and a second round of geometry optimization on each component. A custom Python program, utilizing the OpenBabel library (version 3.1.0),[132] was employed to separate the sucrose from the DBA tweezer and count the number of rotatable bonds (excluding the boronic groups) in the DBA tweezer. This process employed a SMARTS[107] query to identify the sucrose atoms within the complex and restore the hydroxyl groups in both molecules, as shown on the right side of the chemical reaction in Figure 3.7.

DBAs with five or more rotatable bonds were removed from the database, filtering out the most flexible structures. The remaining DBAs underwent a conformational search to estimate the energy difference between the binding conformation for sucrose (E_{Binding}) and the lowest-

energy conformer (E_{Stable}). Ideally, $E_{\text{Binding}} - E_{\text{Stable}} = 0$. This search was carried out using the CREST program (version 2.11)[110] and the semi-empirical GFN2-xTB method,[133] with the GBSA implicit solvent model[114–116] for water. Duplicate DBA structures were eliminated, and the top 100 tweezer-like DBAs, based on the lowest conformational energy differences and the most negative reaction-free energies, were selected for further optimization using an evolutionary algorithm.

3.5 Optimization of Tweezer-Like DBA Molecules with Evolutionary Computational Methods

The tweezer DBA molecules generated by HostDesigner are composed, mainly, of carbon and hydrogen atoms, as the database used for their construction lacks fragments containing heteroatoms or functional groups. Consequently, these tweezers are unlikely to be soluble in aqueous media, necessitating the addition of polar groups to enhance their water solubility. Beyond solubility, the performance of the sucrose-DBA tweezers is also influenced by other critical factors that could be optimized by modifying the initial tweezer's structure, for example, the non-covalent interactions between sucrose and the tweezer molecule, the rigidity and positioning of boronic acid groups, and the pK_a of the PBA moieties. To address these needs, we implemented a modified version of the GB-GA reported by Jensen[66] to perform targeted structural modifications on the sucrose-DBA tweezer molecules produced by HostDesigner.

3.5.1 The GB-GA Evolutionary Method

The GB-GA algorithm utilizes the SMILES[105] string representation to describe molecules as graphs (as shown in Figure 1.15). Consequently, the structural modifications of the DBA tweezers were achieved by modifying the corresponding SMILES strings through genetic operations such as mutation and crossover. The optimization process began by fragmenting the sucrose-DBA complex, as illustrated in Figure 3.8. A core fragment, consisting of sucrose and the boronic acid groups, remained unchanged during GB-GA optimization, while genetic operations were applied to the fragment composed of the linker and the phenyl rings of the aBA moieties. The initial population for the GB-GA algorithm consisted of a database of SMILES strings representing phenyl ring of the PBA molecules and linker fragments from HostDesigner.

A key factor in sucrose selectivity is the precise distance and orientation of the boronic acid

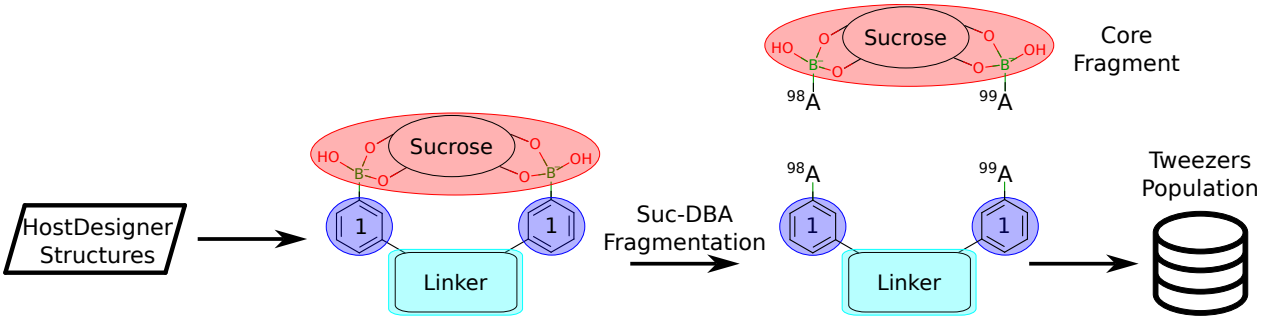


Figure 3.8: Fragmentation of sucrose-DBA complexes for optimization of the tweezer DBA using the GB-GA evolutionary algorithm. The red ellipse highlights the unchanged part of the sucrose-DBA complex during the GB-GA optimization. The light blue and cyan shapes indicate the phenyl rings of the aBA moieties and the linker fragment, respectively, which were modified by the GB-GA algorithm.

groups. To not lose the original connectivity between the core fragment and the DBA tweezer during the genetic modifications, attachment points were labeled with dummy atoms in both the core fragment and the SMILES strings (denoted by "A" in Figure 3.9). Specifically, ^{98}A and ^{99}A represent the attachment sites for the glucose and fructose sides of sucrose, respectively. These attachment points remained unchanged throughout all mutation and crossover operations, ensuring they were conserved in all valid tweezers across generations. Additionally, the Cartesian coordinates of all atoms in the core fragment were preserved to ensure structural accuracy.

Graph modifications of the DBA tweezers through mutation and crossover operations are depicted in Figure 3.9. In case of mutation operation, either one of the aBA moieties or the linker is modified per iteration, with the fragment selected at random for mutation. If no aBA fragments are fused to the linker, the probability of selecting a aBA for mutation is $2/3$, while the probability for the linker is $1/3$. In some cases, one of the aBA sides is fused to the linker, then the probability of selecting either moiety becomes $1/2$. The mutation of a aBA fragment is carried out by randomly replacing the selected aBA with another from the list of aBA molecules found in Appendix B, and the corresponding pK_a value is updated accordingly. These mutations also explore the optimal relative positioning of the boronic group and linker at the $1,2-$ (Table B.1), $1,3-$ (Table B.2), and $1,4-$positions (Table B.3) on the phenyl ring of the aBA, with each substitution being equally probable. Additionally, the selection of aBAs within each substitution position is also equally probable.

The mutations applied to the linker fragment, along with their corresponding probabilities, are illustrated in the right section of Figure 3.10. To expedite the exploration of chemical

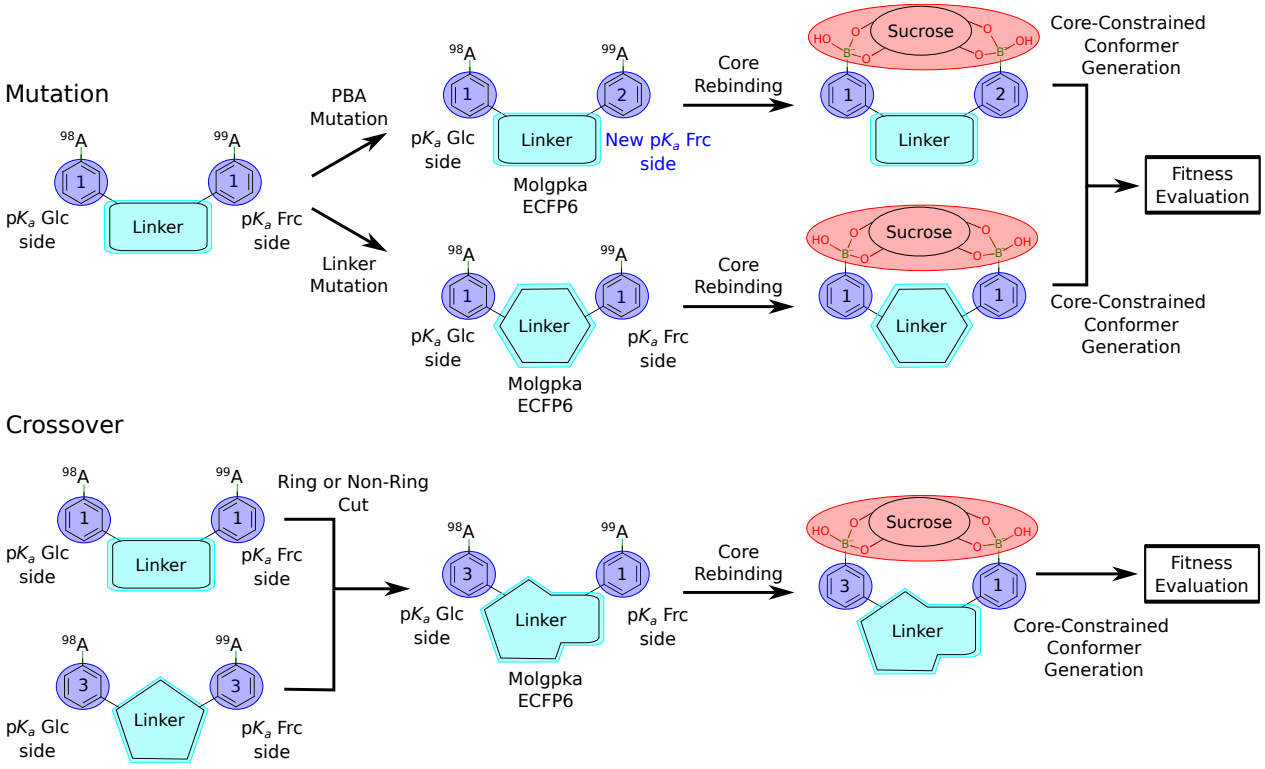


Figure 3.9: Mutation and crossover operations in the GB-GA algorithm for the evolutionary modification of the DBA tweezer moiety. The pK_a values of the aBA fragments were either preserved or updated if the corresponding aBA was modified during the evolutionary process. MolGpka[134] was employed to determine the protonation state of acidic and basic groups in the linker under conditions where pH = max(pK_a) + 1. For aBAs not present in the database (Appendix B), the Extended Connectivity Circular Fingerprints with radius 3 (ECFP6) algorithm[135] was used to predict pK_a based on the structural similarity of the aBA moiety to those in the reported database.

space, we introduced aromatic substitution during the mutation process, where a hydrogen atom bonded to an aromatic carbon is replaced by one of the commonly attached aromatic groups. Unlike the mutation operations implemented by Jensen,[66] which modify one atom or bond at a time and may limit the addition of functional groups to the linker DBA, our approach allows for more flexibility.

The mutations on the linker can introduce basic or acidic functional groups, whose protonation states depend on the solution's pH. When necessary, these protonation states are determined using the MolGpka[134] program, with the reaction condition set to pH = max(pK_a)+1, where max(pK_a) represents the highest value among the aBAs in the tweezer. This ensures that the protonation states of the acid/basic groups in the linker are consistent with the

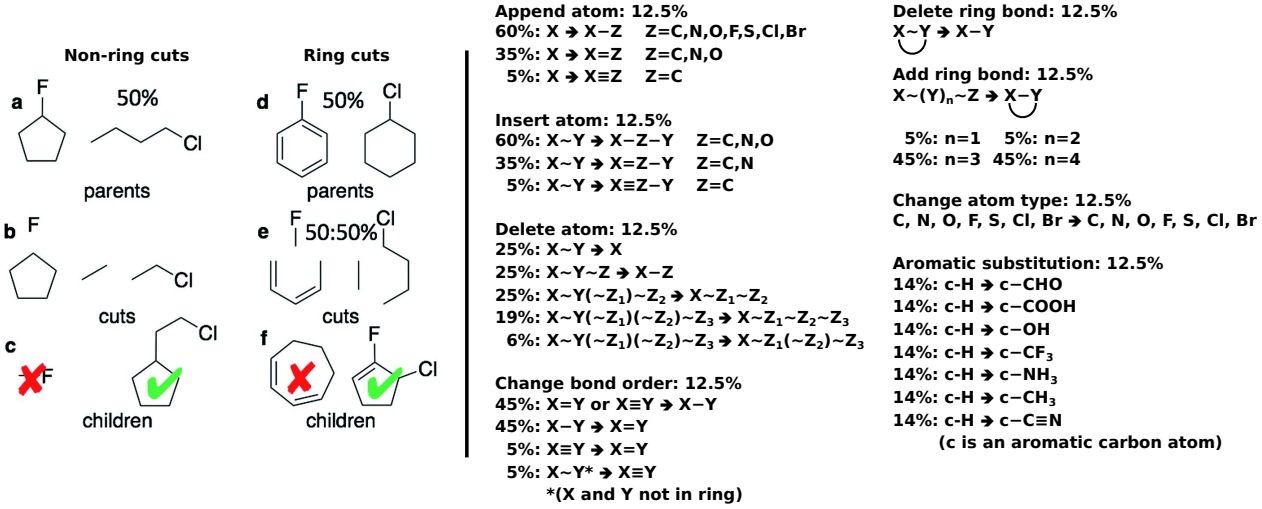


Figure 3.10: (Left) Non-ring and ring cuts of molecules for the crossover operation, and (Right) mutation operations, with their corresponding probabilities, in the GB-GA evolutionary algorithm. The mutation operation is applied exclusively to the linker fragment. An aromatic substitution operation was introduced to incorporate typical aromatic groups into the aromatic linkers during the evolutionary process. Figure modified from reference 66.

tetrahedral boron configuration, which is the most reactive form with carbohydrates[19,45,136] (see Figure 1.3).

The bottom row of Figure 3.9 illustrates the crossover operation for the sucrose DBA tweezers. Two parent molecules were randomly selected from the population, and non-ring or ring cuts were applied to the tweezer structures, as shown on the left side of Figure 3.10. A new DBA tweezer was then generated by fusing the fragments of the parent molecules in a valid configuration. To simplify the implementation, a mutation operation was applied post-crossover to restore the pK_a values of the aBA moieties. In cases where the crossover produced novel or unknown aBA moieties, the Extended Connectivity Circular Fingerprints with radius 3 (ECFP6) algorithm[135] was employed to generate a bitwise representation of the fragmented aBA. The pK_a value was assigned based on the closest similarity, calculated as the average of Dice, Tanimoto, and Cosine similarity scores,[137] using a non-redundant list the aBA molecules in Appendix B.

After performing the crossover and mutation operations, the chemical validity of the newly generated tweezers was assessed. This included checking atom valences, bond orders, the presence of attachment points for sucrose and boronic groups, ring sizes, and aromaticity. Tweezers containing more than 40 atoms were discarded.

3.5.2 The Multi-Objective Optimization Function

The GB-GA algorithm iteratively optimized the DBA tweezers through multi-objective minimization problem, as defined in Equation 3.5

$$\min[J(\Delta_r G^\circ, \log P_{\mathrm{o/w}}, \max(\mathrm{p}K_a), \text{Synthetic Accesibility (SA)}]. \tag{3.5}$$

The fitness of the chemically valid tweezers was evaluated using the score function in Equation (3.6), and they were ranked in ascending order according to their values of J. Equation (3.6) considered four objective terms: the reaction free energy $\Delta_r G^\circ$ (as defined by Equation (1.30)), the octanol-water partition coefficient ($\log P_{\mathrm{o/w}}$), the maximum $\mathrm{p}K_{a,\mathrm{aBA}}$ value of the aBA fragments, and a SA score was also included in the fitness function to simplify the synthetic feasibility of the DBA tweezers. The score value J for each individual in the tweezers population is a normalized weighted sum of the four components:

$$\begin{aligned} J = c_1 \cdot \Delta_r G^\circ + c_2 \cdot \log P_{\mathrm{o/w}} + c_3 \cdot \max(\mathrm{p}K_{a,\mathrm{aBA}}) + c_4 \cdot \mathrm{SA}; \\ \sum_{i=1}^{i=4} c_i = 1. \end{aligned} \tag{3.6}$$

The terms in Equation (3.6) were translated to the same scale by a MinMax scaling according to the following range of values:

$$\begin{aligned} x_{\mathrm{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} &\in [0, 1]\,; \\ \Delta_r G^\circ &\in [-30\,\mathrm{kcal/mol}, 1\,\mathrm{kcal/mol}]\,; \\ \log P_{\mathrm{o/w}} &\in [-3, 6]\,; \\ \max(\mathrm{p}K_{a,\mathrm{aBA}}) &\in [3, 9]\,; \\ \mathrm{SA} &\in [1, 10]\,. \end{aligned} \tag{3.7}$$

To improve the efficiency of the evolutionary optimization of the DBA tweezers using the GB-GA algorithm, it was essential to reduce the computational cost of evaluating $\Delta_r G^\circ$ in the multi-objective Equation 3.6. For screening purposes, the relative differences in $\Delta_r G^\circ$ between the DBA tweezers are more critical than obtaining precise absolute values. Additionally, the screening process benefits from a reduced computational cost, enhancing throughput. The details for calculating $\Delta_r G^\circ$ during the evolutionary process are provided in Section 4.5.

The octanol-water partition coefficient, $P_{\mathrm{o/w}}$ (Equation (1.31)), was calculated for the

neutral DBA tweezer. Geometry optimizations were performed using the GFN2-xTB [133] semiempirical method, with the ALPB implicit solvent model [113] for both water and octanol, as implemented in the xtb program suite. [138] This was used to compute the solvation free energies, $\Delta G^{\circ}_{\mathrm{solv,o}}$ and $\Delta G^{\circ}_{\mathrm{solv,w}}$. The maximum $\mathrm{p}K_{a,\mathrm{aBA}}$ value, $\max(\mathrm{p}K_{a,\mathrm{aBA}})$, was directly derived from the tweezer's structural information, as explained in the previous section. Finally, the SA score was computed using the algorithm reported by Ertl and Schuffenhauer. [139]

The selection of individuals for crossover or mutation was carried out randomly, with probabilities (p_i) determined by linear ranking selection: [140]

$$p_i = \frac{1}{N}\left(\eta - (2\eta - 2)\frac{i-1}{N-1}\right), \tag{3.8}$$

where i is the tweezer's ranking, N is the population size, and η is the selection pressure with values ranging from $1 \leq \eta \leq 2$. When $\eta = 1$, all individuals have an equal likelihood of being chosen for genetic operations, promoting exploration of the chemical space. In contrast, $\eta = 2$ favors the selection of the fittest individuals by assigning probabilities based on rank, thereby driving the optimization process towards more favorable solutions.

The probabilities for crossover and mutation operations, along with the selection pressure parameter η, were optimized to improve the efficiency of the genetic algorithm. The entire implementation described in this section was developed in Python, utilizing the RDKit library [106] for graph-based representations of the tweezers, encoded via SMILES notation. [105] The RDKit library was also employed to perform the genetic operations. Figure 3.11 illustrates the overall methodology used for designing the DBA tweezer for selective sucrose detection in aqueous media.

3.6 Synthesis of Tweezer-like DBA Molecules

Developing efficient synthesis methods for polyaromatic compounds has enabled their use in various applications, including molecular electronics, photovoltaic dyes, and sensors. [141] Continuos improvements on the performance of Suzuki-Miyaura reaction, alkylation of activated carbonyl compounds, [141] and inter- and intramolecular metal-catalyzed cross-coupling reactions, [142] have been instrumental in designing tweezer-like DBA molecules. These methods offer several advantages over traditional techniques, such as improved reaction yields, fewer reaction steps, enhanced control over the geometry and extension of the aromatic system, and milder reaction conditions. [141] However, challenges remain, particularly in optimizing re-

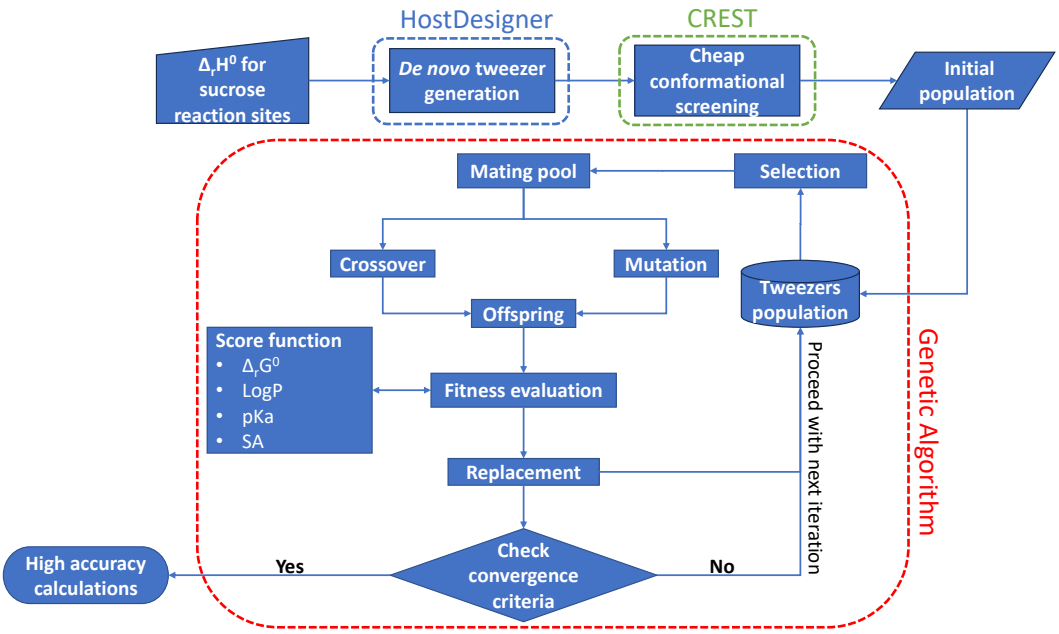


Figure 3.11: Flowchart illustrating the *in silico* design of DBA tweezer-like molecules for the selective detection of sucrose. Dashed lines indicate the steps executed by HostDesigner, CREST, and the GB-GA algorithm.

action yields based on the type and position of functional groups. This continues to be an area of active research in polyaromatic compound synthesis.[141]

In this work, we utilized Pd-catalyzed cross-coupling reactions as the basis for the synthesis of one of the tweezer-like DBA molecules designed for sucrose binding. This process involved constructing strained polyaromatic systems and precisely positioning boronic groups within these systems to enable effective sucrose binding. The synthesis results are presented in Section 4.6.

3.7 Simulation of SAM Coating an Au Surfaces

We performed MD simulations to study the interface between liquid water and a SAM coating on an Au(111) surface. The goal was to determine the surface coverage of a sucrose-binding DBA tweezer in an aqueous environment. The SAM coating was modeled as a combination of 4-Mercaptobenzoic acid (MBA) capped with methylamine ($CH_3{-}NH_2$) and one DBA tweezer designed for sucrose binding. The 4-MBA molecules were bonded to the Au(111) surface, arranged in a $\left(\sqrt{3} \times 4\right)$ surface lattice.[143,144]

The system was assembled using the Avogadro program (version 1.99.0)[145] for slab con-

struction. A custom Python script, utilizing the OpenBabel library (version 3.1.0),[132] was employed to position the thiol molecules and form the SAM coating.

MD simulations were conducted using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software package.[146] The CHARMM force field was used to model the system, incorporating harmonic potentials for bonds and angles, CHARMM-style dihedral potentials, and harmonic improper potentials. Non-bonded interactions were described using the Lennard-Jones potential with a cutoff of 12 Å, and long-range Coulombic interactions were treated using the Particle-Particle Particle-Mesh (PPPM) method with a precision of 1×10^{-6}. The SHAKE algorithm was applied to constrain bond angles involving hydrogen atoms, enabling a timestep of 2 fs.

The production run was performed under isothermal-isobaric (NPT) ensemble conditions at a temperature of 303.15 K and a pressure of 1 bar. Temperature and pressure were controlled using a Nosé-Hoover thermostat and barostat, with the system coupled in all three dimensions (xyz) to maintain isotropic pressure. The simulation was carried out for 20 ns.

Input files and force field parameters (Section 1.2.2.3) for the MD simulations of the water-SAM-Au(111) interface were prepared using the CHARMM-GUI web interface,[147] Specifically the modules for PDB reading[148–150] and multicomponent assembly.[151] Missing force field parameters for 4-MBA capped with $CH_3{-}NH_2$ and the DBA tweezer for sucrose were obtained using the CHARMM General Force Field (CGenFF) web service.[152]

Chapter 4

Results and Discussion

The selection of DBA tweezers for developing selective sucrose sensors in this work was based on their proven advantages over other molecular recognition methods for carbohydrates:[12] (i) DBA tweezers has demonstrated its biocompatibility and operability for quantification of glucose under *in vivo* conditions. (ii) Compared to enzyme proteins such as INV and SuSy used in biosensors, DBA tweezers are low-molecular-weight molecules, making them more adaptable to different operational conditions. (iii) DBA tweezers exhibit versatility in signal transduction for carbohydrate quantification, supporting their implementation in sensing platforms based on, for example, fluorescence, spectroscopy, or electrochemical methods. (iv) The modular synthesis approach for DBA tweezers designed for glucose can significantly streamline their adaptation for sucrose.

Despite the numerous benefits of DBA tweezers for carbohydrate quantification, there are currently no reported examples of DBA tweezers specifically designed for sucrose. In our view, the inverse design strategy proposed by Yang and coworkers[17] provides a promising starting point for the computational aided design of DBA tweezers for sucrose. However, Yang's methodology[17] has notable limitations, as it relies solely on a structure-based approach and lacks the incorporation of physicochemical parameters necessary to ensure optimal performance of the designed DBA tweezers under physiological conditions. To address these gaps, we extended the inverse design strategy by integrating physicochemical parameters and a SA score into the computational design process of DBA tweezers.

4.1 Calculation of Free Energies for Reaction of DBA Molecules with Monosacharides

The reaction-free energy ($\Delta_r G^\circ$) is an important physicochemical parameter considered in the design of DBA tweezers for sucrose. Negative values of $\Delta_r G^\circ$ indicate that the products of a reaction are more stable than the reactants, resulting in energy being released into the environment, while positive values suggest that the products are less stable than the reactans, requiring energy from the environment to form the products under standard conditions (indicated by the "$^\circ$" symbol).

Calculating reaction-free energies for DBA tweezers and carbohydrates in aqueous media poses challenges because it is essential to consider solute-solvent interactions, changes in conformational entropy, and conformer ensemble averaging in order to obtain accurate values. To tackle these challenges, we developed a semiautomatic workflow[108] designed for these calculations and demonstrated its application by comparing the results with the experimental $\Delta_r G^\circ$ values for the reactions of tweezers **DBA-1**$^{-2}$ and **DBA-2** with monosaccharides Glucose (Glc) and Galactose (Gal), as shown in Figures 3.1, 3.2, and 3.3.

The following section provides details on the development, fine-tuning, and results of the free energy calculations for the reaction of DBA tweezers with monosaccharides.

4.1.1 Conformational Entropy Contribution

Given the substantial reduction in molecular flexibility from the reactants (**DBA-1**$^{-2}$, **DBA-2**, Glc, and Gal) to the bound complexes (**DBA-1.Glc**$^{-2}$, **DBA-2.Glc**, and **DBA-1.Gal**$^{-2}$), the contribution of conformational entropy to the reaction-free energy is likely to be significant and should be taken into account. To investigate the impact of entropic penalties in our specific set of systems, the entropic contribution $-T\Delta S^\circ_{\text{conf}}$ is calculated from the change in entropy $\Delta S^\circ_{\text{conf}}$ from reactants to products at 298.15 K. The statistical noise in the ensemble generation with CREST is taken into account by averaging over three calculations for each system.

The entropic contribution to the free energy for the **DBA-1**$^{-2}$+Glc, **DBA-2**+Glc, and for **DBA-1**$^{-2}$+Gal reaction are presented in Table 4.1. The negative $T\Delta S^\circ_{\text{conf}}$ values qualitatively agree with the closure of two rings via the formation of four boronate ester (B−O−C) bonds, depicted in the Figures 3.1, 3.2, and 3.3, and the locks of the dihedral rotations of the PBA groups. Additionally, the reduction in conformational entropy $T\Delta S^\circ_{\text{conf}}$ is compara-

tively lower for **DBA-2**+Glc than for **DBA-1**$^{-2}$+Glc and **DBA-1**$^{-2}$+Gal reactions. These differences arise because, in **DBA-2**, the rotations of the PBA groups are dampened by the anthracene ring, lowering the S°_{conf} for **DBA-2** compared with **DBA-1**$^{-2}$, in which the PBA groups are freer to rotate. These $T\Delta S^{\circ}_{\text{conf}}$ values introduce a shift of the reaction-free

Table 4.1: Calculated conformational entropy contribution for the reaction leading to the given complex. The four water molecules produced after the reaction were removed from the reaction equations because they did not contribute to the conformational entropy change.

Reaction	$T\Delta S^{\circ}_{\text{conf}}$ [kcal/mol]
DBA-1^{-2} + Glc $\rightleftharpoons$ **DBA-1.Glc**$^{-2}$	-4.82
DBA-2 + Glc $\rightleftharpoons$ **DBA-2.Glc**	-3.31
DBA-1^{-2} + Gal $\rightleftharpoons$ **DBA-1.Gal**$^{-2}$	-4.11

energies $\Delta_r G^{\circ}_{\text{CE}}$, significantly improving the accuracy (see below). Generally, it is advisable to analyze how the number of relevant conformers changes between the reactants and the products. Practically, the number of conformers needed to recover 90 % of the Boltzmann weight can be used as a measure of flexibility and can be obtained from a CREST calculation. If this measure changes significantly during the reaction, the conformational entropy should be considered in the free energy calculation, as is exemplified by the reactions reported in this work. Comparing the number of relevant conformers for the Boltzmann average after the CENSO refinement, the product conformers decrease by approximately a factor of ∼12 compared with the reactant conformers, no matter the level of theory used for the final refinement with CENSO. These results are reported in the first rows ($(H_2O)_0$) of Tables A.4 and A.6 in the Appendix A. Hereafter, the reported reaction-free energies $\Delta_r G^{\circ}_{\text{CE}}$ (including those in the Appendix A) incorporate the corresponding conformational entropy changes $T\Delta S^{\circ}_{\text{conf}}$ from Table 4.1, as given by equation (1.30).

4.1.2 Reaction-Free Energies ($\Delta_r G^{\circ}$) with the Implicit Solvation Model

Implicit solvation models offer an efficient means of capturing the interactions between solute and solvent molecules, making them a preferred choice for calculating the physicochemical properties of molecules and reactions in solvent environments. In Table 4.2, we present the free energies $\Delta_r G^{\circ}_{\text{CE}}$ for the reactions of **DBA-1**$^{-2}$+Glc and **DBA-2**+Glc employing the SMD implicit solvent model and different levels of theory. Irrespective of the level of theory, the computed $\Delta_r G^{\circ}_{\text{CE}}$ values for both reactions significantly overestimate the corresponding experimental references, by approximately a factor of 2 for **DBA-1**$^{-2}$+Glc and a factor of 3

Table 4.2: Reaction-free energies (kcal/mol) for the reactions **DBA-1**$^{-2}$+Glc and **DBA-2**+Glc calculated with the SMD model and different levels of theory. The energies are calculated from the median (equation 3.2) of a distribution of reaction-free energies $\{\Delta_r G^\circ_{\mathrm{CE},x}\}$ generated with three trials (N = 3) values of molecular free energies $G^\circ_{\mathrm{CE},i}(X)$.

Level of theory	**DBA-1**$^{-2}$+Glc	**DBA-2**+Glc
B97-D3	-12.6 ± 0.1	-17.0 ± 0.2
r^{2}SCAN-3c	-9.2 ± 0.1	-13.3 ± 0.2
PW6B95-D4/def2-TZVPD[a]	-12.5 ± 0.4	-16.0 ± 0.1
ωB97X-D4/def2-TZVPD[a]	-11.0 ± 0.1	-15.0 ± 0.1
Experiment	-5.8 [53,120]	-4.6 ± 0.1 [55] -5.5 [55]

[a]The molecular free energies $G^\circ_{\mathrm{CE}}(X)$ were calculated over one of the CE optimized at level of theory r^{2}SCAN-3c

for **DBA-2**+Glc. This discrepancy suggests a systematic error source in the computation of the reaction-free energies $\Delta_r G^\circ_{\mathrm{CE}}$.

Figure 4.1 depicts representative structures of the energetically most favorable conformers of **DBA-1**$^{-2}$ and **DBA-2**, as obtained from censo refinement of the CE with setting B, and the implicit solvation SMD. Both molecules exhibit intramolecular hydrogen bonds. In **DBA-1**$^{-2}$, these form between the B−O−H and methoxy groups (see Figure 4.1a) and are present in an energy range of, approximately, 0.5 kcal/mol relative to the lowest conformer. In **DBA-2**, the intramolecular hydrogen bonds are located between the B−O−H groups, as shown in Figure 4.1b, and are present in an energy range of 2.0 kcal/mol relative to the lowest conformer. We attribute the overstabilization of the given conformers to the inability of implicit solvent models to accurately represent the strong interactions between the solute and solvent molecules,[111,153–156] particularly when molecules feature highly polarized groups such as the boronic groups in the DBA reactants. Forming these hydrogen bonds may also induce molecular stress in the geometry of the reactants **DBA-1**$^{-2}$ and **DBA-2**, increasing their molecular energies $G^\circ_{\mathrm{CE}}(X)$. Additionally, implicit solvation models struggle to calculate the solvation-free energy for anionic molecules[157,158] because estimating the solvent-accessible surface is difficult due to the potentially incorrect radii used for the cavity construction. The poor radii are discussed as a reason for the poor performance of implicit solvation models for negatively charged molecules.[157,159] This problem has recently been addressed by the dynamic radii adjustment for continuum solvation (DRACO).[158]

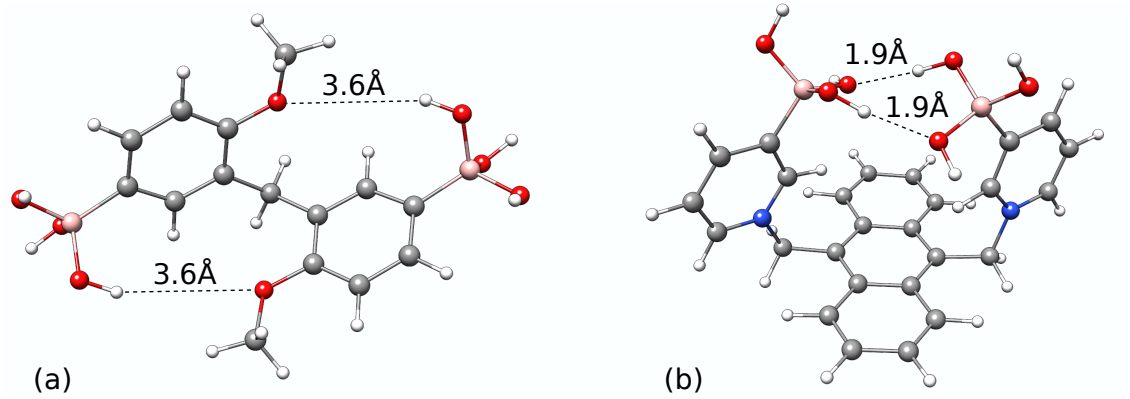


Figure 4.1: Stable conformers of **DBA-1**$^{-2}$ (a) and **DBA-2** (b) molecules in implicit solvent. The dashed lines are a visual aid for the intramolecular hydrogen bond interactions.

A visual inspection of the CE for glucose does not reveal any of the above-mentioned problems for **DBA-1**$^{-2}$ and **DBA-2**. Moreover, Feng et al.[160] showed that HF/6-31+G(d) level of theory with the implicit solvation model C-PCM is already accurate enough for the calculation of acidity constants pK_a and proton affinities for glucose.

To improve the representation of the aqueous media in our calculations, we decided to add explicit water molecules around the solute molecules in a cluster-continuum approach. The results of the first phase of the parameter search are presented in the next section.

4.1.3 Screening of CENSO Parameters for Microsolvation

In the previous section, the average computation time per conformer for the calculation of CENSO's part 2 of a CE of **DBA-1**$^{-2}$.Glc is half of the time compared to **DBA-2**.Glc. Therefore, we have selected the reaction **DBA-1**$^{-2}$+Glc for the calculations in this section to save computation time. The number of water molecules for microsolvation was set arbitrarily to seven water molecules for the free ligand **DBA-1**$^{-2}$ and glucose respectively, to provide a first insight into the impact of microsolvation. The complex **DBA-1.Glc**$^{-2}$ is calculated with six water molecules even though four hydroxyl groups are part of the boron ester bonds, to avoid big differences in the number of explicit solvent molecules compared to the reactants. As the distributions sampled in the three trials might not follow a normal distribution, we report median values and IQR hereafter. Histograms of the sampled energy distributions $\{\Delta_r G^\circ_{\mathrm{CE},x}\}$ are reported in Figure A.6.

The results in Table 4.3 demonstrate that incorporating explicit water molecules for all the reactants and products, in a microsolvation approach with the QCG workflow, enables more

Table 4.3: Reaction-free energies $\Delta_r G^\circ_{\mathrm{CE}}$, IQR, and deviation of from experiment $\Delta\Delta G^\circ_{\mathrm{CE\text{-}CD}}$ for the reaction **DBA-1**$^{-2}$ + Glc with the CENSO settings listed in Table 3.2. All values are reported in kcal/mol.

Set.	$\Delta_r G^\circ_{\mathrm{CE}}$	IQR	$\Delta\Delta_r G^\circ_{\mathrm{CE\text{-}CD}}$
E	-6.7	2.6	-0.9
F	-2.9	1.6	2.9
C	-2.7	4.1	3.1
G	-2.2	2.3	3.6
B	-2.0	1.7	3.8
A	0.1	7.0	5.9
D	-19.3	2.0	-13.5
Experiment	-5.8	-	

The sets of $\{\Delta_r G^\circ_{\mathrm{CE},x}\}$ and the IQRs for the **DBA-1**$^{-2}$ + Glc reaction were calculated with the eqs 3.2, (A.5), and (A.6), using three trials for both $\{G^\circ_{\mathrm{CE},\alpha}(X(\mathrm{H_2O})_n)\}$ and $\{G^\circ_{\mathrm{CE},\beta}((\mathrm{H_2O})_n)\}$.

accurate calculations of free energies for the reaction **DBA-1**$^{-2}$ + Glc $\overset{K}{\rightleftharpoons}$ **DBA-1.Glc**$^{-2}$ + 4 $\mathrm{H_2O}$. The systematic error in $\Delta_r G^\circ_{\mathrm{CE}}$ is reduced from around 6 kcal/mol (with only the implicit solvation model) to approximately 3 kcal/mol. The only exceptions are settings A and D, which still show significant deviations from the experimental reference.

Calculations with settings A, B, and C were performed to evaluate the influence of energy thresholds G_i^{thr} used to sort out the high-lying energy conformers with CENSO on the free energies G°. The Kruskal-Wallis test[161] and Dunn's test indicate significant differences between settings A, B, and C concerning the calculated reaction-free energies. Further details are provided in the Supporting Information in Table A.1. These results imply an important influence of the CENSO energy thresholds G_i^{thr} in the distribution of the values $\{\Delta_r G^\circ_{\mathrm{CE},x}\}$. Nevertheless, the difference in median values between settings B and C is less than 1.0 kcal/mol meaning that, with settings B, the differences in the PES between GFN2-xTB (used for the CE sampling) and the DFA methods used in the CENSO parts are already taken into account to an acceptable amount. However, setting B is, on average, 22 times faster than setting C for refining a CE of **DBA-1**$^{-2}$.Glc, highlighting the significant impact of proper tuning of G_i^{thr} on the computational efficiency of the workflow. Meanwhile, the difference in median values of settings A and B highlights the importance of CENSO energy thresholds G_i^{thr} as crucial parameters for the accuracy of $\Delta_r G^\circ_{\mathrm{CE}}$. The current default thresholds of $G_0^{\mathrm{thr}} = 4.0$ kcal/mol, $G_1^{\mathrm{thr}} = 3.5$ kcal/mol, and $G_2^{\mathrm{thr}} = 2.5$ kcal/mol recommend a careful evaluation of the optimal values for these parameters before performing a large-scale calcu-

lation for molecular free energies $G^{\circ}_{\mathrm{CE}}(X)$ of similar systems that require microsolvation due to strong solute-solvent interactions.

Settings C, D, E, F, and G from Table 3.2 were selected to test the accuracy of different combinations of DFA methods for calculating $\Delta_r G^{\circ}_{\mathrm{CE}}$. While setting F gives similar results compared with settings C and G ($|\Delta\Delta_r G^{\circ}_{\mathrm{CE\text{-}CD}}| \sim 3\,\mathrm{kcal/mol}$), setting D gives the worst results among the settings tested ($|\Delta\Delta_r G^{\circ}_{\mathrm{CE\text{-}CD}}| > 10\,\mathrm{kcal/mol}$). The lack of accuracy in setting D likely stems from the electronic energy $\overline{E}_{\mathrm{gas}}(X)$ and solvation free energy $\overline{\delta G}_{\mathrm{solv}}(X)$ with the HF-3c/MINIX method, which, due to the minimal basis set, fails to accurately capture polarization effects,[162] crucial for representing solute-solvent interactions and B−O bond polarization in DBA molecules. Therefore, setting D was excluded from further analysis.

For settings C, E, F, and G, Dunn's test showed significant differences ($p < 0.05$) between the medians values for $\Delta_r G^{\circ}_{\mathrm{CE}}$. However, the median values for settings C, F, and G are in close agreement within a range of less than 1 kcal/mol, with an absolute error of around 3 kcal/mol to the experimental reference $\Delta_r G^{\circ}_{\mathrm{CD}}$ for the reaction of reaction **DBA-1**$^{-2}$+Glc. Moreover, setting E gives the most accurate result with an absolute error of 0.9 kcal/mol compared with the experimental value. For all these settings, the final molecular free energy G°_{CE} is calculated with a TZ basis set, which, along with microsolvation, explains the agreement with the experimental reaction-free energy. Nevertheless, in the case of setting E, we do not discard a fortuitous (passive) error cancellation as the main reason for the improved accuracy because the only difference between settings E and G is the level of theory used in part 0 (cheap pre-screening) which, according to the results with the other settings, should not considerably improve the calculated $\Delta_r G^{\circ}_{\mathrm{CE}}$.

When the given workflows are applied for screening purposes, a small IQR becomes more important, since the ranking of the target property is ideally not influenced by the accuracy of the method but by its precision. In this regard, settings B and F are the most favorable with an IQR of 1.7 kcal/mol and 1.6 kcal/mol respectively. Furthermore, a robust or precise workflow, with the optimal tuning of energy threshold G_i^{thr}, allows the use of fewer trials, thus saving valuable computational costs. When comparing IQRs to define the precision in this way, it is important to ensure that the underlying distributions are unimodal and free of outliers. Considering the accuracy, precision, and computational costs of the workflows discussed in this section, the best-performers are settings B, E, F, and G.

4.1.4 Screening for the Optimal Number of Explicit Water Molecules

Since microsolvation is imperative to obtain accurate reaction-free energies for the DBAs, the number of explicit water molecules $(H_2O)_n$ is screened using the best-performing settings from the previous section (B, E, F, and G). To save computational time, the energy thresholds G_i^{thr} are reduced for all workflows to those used in setting B. All reactants and products were microsolvated with the same number of water molecules $(H_2O)_n$, and the energy distributions $\{\Delta_r G^\circ_{\mathrm{CE},x}\}$ were calculated using the molecular free energies $G^\circ_{\mathrm{CE},i}(X)$ from three trials. To avoid confusion with the previous configurations, in this section we renamed the settings E, F, and G as E2, F2, and G2, respectively.

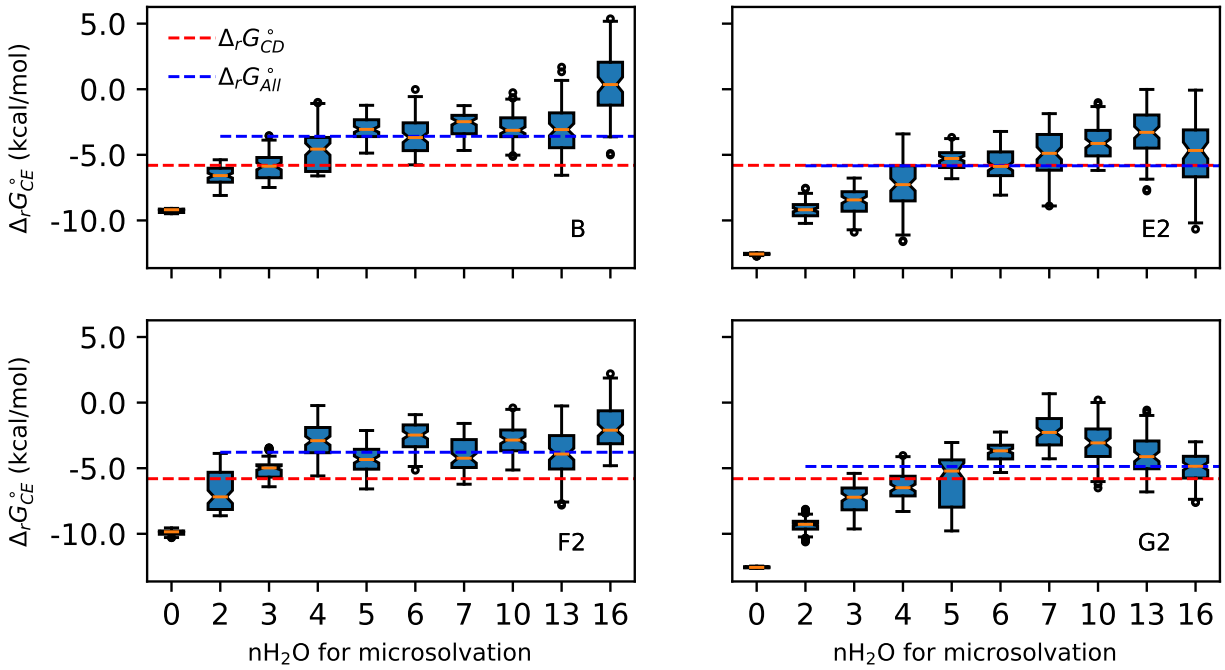


Figure 4.2: BoxPlot diagrams for the reaction-free energies ($\Delta_r G^\circ_{\mathrm{CE}}$) of **DBA-1**$^{-2}$ and glucose, with different numbers of water molecules for microsolvation. The dashed red line serves as a visual guide for the experimental reaction-free energy ($\Delta_r G^\circ_{\mathrm{CE}}$) calculated from CD experiments.[53,120]

Figures 4.2 and 4.3 show the energy distributions as boxplots for the reactions **DBA-1**$^{-2}$+Glc and **DBA-2**+Glc, respectively. Tables A.3 and A.5 show the $\Delta_r G^\circ_{\mathrm{CE}}$, IQR, and signed error $\Delta\Delta_r G^\circ_{\mathrm{CE\text{-}Exp}}$ for Figures 4.2 and 4.3, respectively. Introducing explicit water molecules improves the representation of solute-solvent interactions, but also increases the IQR of $\Delta_r G^\circ_{\mathrm{CE}}$ from less than 0.5 kcal/mol to over 3 kcal/mol, highly reducing the precision

in some cases. Furthermore, using CENSO's part 2 of the **DBA-1**$^{-2}$**.Glc** product as a case example, the computational time per conformer for setting B increases proportionally with the number of water molecules used in microsolvation compared to the implicit solvent model ($n = 0$). The decrease in precision and the rise in computational cost are both expected, as the system's degrees of freedom increase with the number of explicit solvent molecules, necessitating more extensive conformational space sampling. Overall, the fluctuations of the reaction-free energies concerning the number of explicit water molecules lie around 8 kcal/mol irrespective of the DBA or applied method. Although there is a number of solvent molecules for each setting that reproduces the experimental reaction-free energies with high accuracy, there are no clear criteria in the literature to define when a system is correctly microsolvated. Matching the calculated $\Delta_r G^\circ_{\mathrm{CE}}$ and experimental $\Delta_r G^\circ_{\mathrm{Exp}}$ is not feasible for predictive workflows where experimental data is unavailable. Furthermore, different numbers of explicit solvation molecules can lead to matches in reaction-free energies, as observed for **DBA-1**$^{-2}$ with setting G2.

In principle, the double zwitterionic nature of **DBA-2** and **DBA-2.Glc** at pH 7.4 makes these molecules more polar, necessitating a greater number of explicit water molecules to accurately represent the solute-solvent interactions.

Other criteria for considering a solute to be correctly microsolvated may be related to structural features. Ideally, the optimal number of solvent molecules would correlate with a relaxation of a bond angle or dihedral angle in the reactants.

However, based on the investigated angles given in Tables A.3 and A.5, it is not possible to decide on one specific number of solvent molecules in our case. The main problem is that the fluctuations in the bond angle when looking at different conformers, changing the number of solvent molecules, or changing the applied method are as large as the observed relaxation. In the absence of a universal criterion, we propose to use the median of all reaction-free energies that include microsolvation $\Delta_r G^\circ_{\mathrm{All}}$ for each of the settings, to provide a single value for each setting. For **DBA-1**$^{-2}$+Glc the final energy values deviate from the experimental values $\Delta_r G^\circ_{\mathrm{CD}}$ by 2.2, -3.5×10^{-2}, 2.0, and 0.9 kcal/mol for settings B, E2, F2, and G2, respectively. For **DBA-2**+Glc, the calculated energies for settings B, E2, F2, and G2 deviate by -3.1, -4.6, -3.1, and -4.3 kcal/mol from the experimental value $\Delta_r G^\circ_{\mathrm{PT}}$, and by -2.2, -3.7, -2.2, and -3.4 kcal/mol from the experimental value $\Delta_r G^\circ_{\mathrm{NMR}}$. With the given precision and accuracy, our workflow can be used for screening candidate DBAs when using a reasonably large threshold.

In the following, further findings on the application of microsolvation are discussed. The

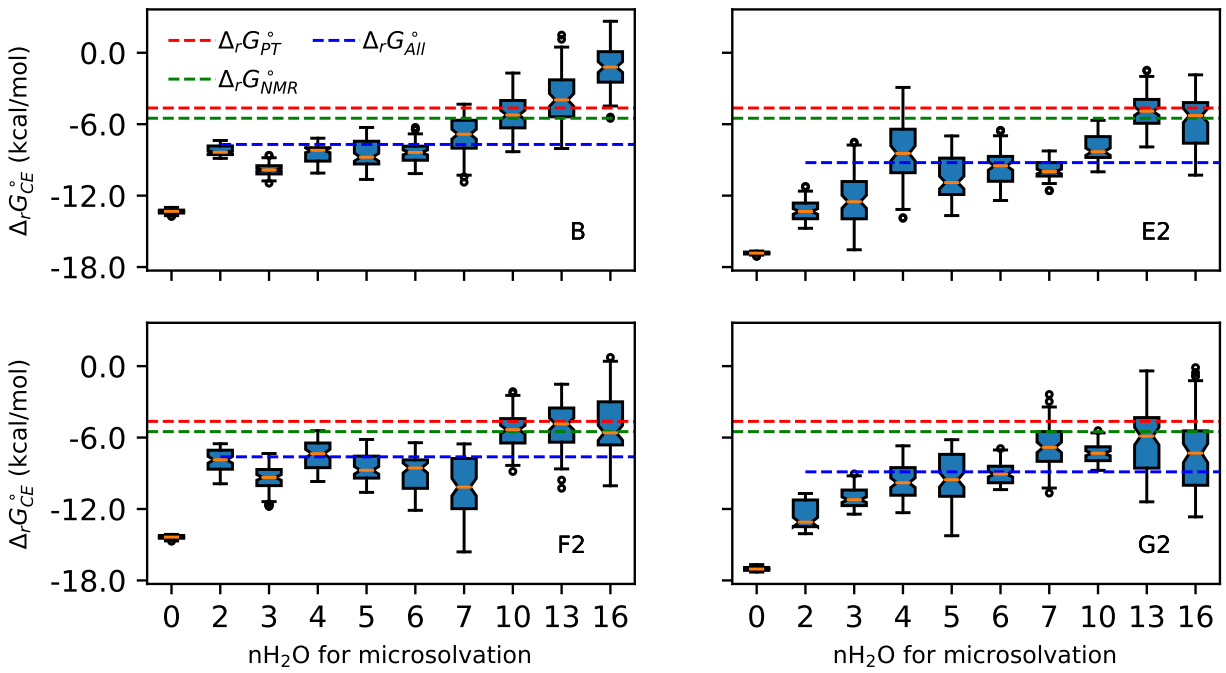


Figure 4.3: BoxPlot diagrams for the reaction-free energies ($\Delta_r G^\circ_{\mathrm{CE}}$) of **DBA-2** and glucose, with different numbers of water molecules for microsolvation. The dashed red and green lines serves as a visual guide for the experimental reaction-free energies ($\Delta_r G^\circ_{\mathrm{PT}}$ and $\Delta_r G^\circ_{\mathrm{NMR}}$) calculated from PT and NMR experiments, respectively.[55]

RMSD from the initial configuration (Figures A.2, A.4, A.8, A.10, and A.13) decreases when increasing the number of water molecules for microsolvation, while the time constants (T_x) for the RMSD's auto-correlation functions (Figures A.3, A.5, A.9, A.11, and A.14) become greater. This indicates that a strong network of hydrogen bonds formed by the water molecules[155] reduces the conformational changes of the solutes during the 80 ps of metadynamics sampling with CREST. As a consequence, the number of relevant conformers in the Boltzmann average for $G^\circ_{\mathrm{CE}}(X)$ also decreases when adding water molecules to the solute, as reported in Tables A.4, A.6, and A.7. Simm et al. also reported a decrease in relevant conformers as the number of water molecules increased in the microsolvation of acetonitrile with water.[155] Therefore, this trend is not related to any numerical artifact in our workflow. Additionally, it has been shown that random error propagation increases when few microstates (or conformers) are used in the calculation of statistical-derived thermodynamic properties[163] (like the molecular free energies $G^\circ_{\mathrm{CE}}(X)$ and reaction-free energies $\Delta_r G^\circ_{\mathrm{CE}}$), which overall agrees with the IQR trends in Figures 4.2 and 4.3, and Tables A.3 and A.5. Considering the

above-mentioned arguments, we can not discard the matching between the calculated $\Delta_r G^\circ_{\text{CE}}$ and experimental $\Delta_r G^\circ_{\text{Exp}}$, as a consequence of poor sampling of the configuration space or incorrect selection of the DFA methods for the CE refinement, being an example of "get the correct answer by the wrong reasons". We conducted pairwise Dunn's tests to assess the convergence of calculated reaction-free energies depending on the number of water molecules used for microsolvation. Heatmaps for the pairwise Dunn's tests are reported in Figures A.7 and A.12 for reactions the reactions **DBA-1**$^{-2}$+Glc and **DBA-2**+Glc, respectively. In general, for **DBA-1**$^{-2}$+Glc reaction, the optimal number of water molecules for microsolvation is in a range between 5-10 molecules because Dunn's test does not show significant differences between the $\Delta_r G^\circ_{\text{CE}}$ in this range. In the case of **DBA-2**+Glc reaction, Dunn's test does not show significant differences for $\Delta_r G^\circ_{\text{CE}}$ in two ranges, a first range of 4-6 water molecules where the calculated amount of energy released by the reaction is overestimated compared to the experimental values, and a second range from 10 to 16 water molecules where the reaction-free energy for **DBA-2**+Glc match with experimental references $\Delta_r G^\circ_{\text{PT}}$ and $\Delta_r G^\circ_{\text{NMR}}$ in most of the settings. However, pairwise Dunn's tests may require more trials N for $G^\circ_{\text{CE}}(X)$ to achieve conclusive results regarding the convergence of $\Delta_r G^\circ_{\text{CE}}$, with the number of water molecules for microsolvation. Due to the computational cost, it may be worth exploring this criterion in other systems with smaller conformational space for the solute.

The reliability of the results presented in Figures 4.2 and 4.3, and the optimal parameters for microsolvation, are validated in the next section, by using more trials for the calculation of reaction-free energies.

4.1.5 Workflow Accuracy

All the previous calculations of the reaction-free energy $\Delta_r G^\circ_{\text{CE}}$ were performed with three trials. To test the accuracy of our workflow, we increased the number of samples for $G^\circ_{\text{CE},i}(X)$ up to ten trials ($N = 10$) for the calculation of the energy distributions $\{\Delta_r G^\circ_{\text{CE},x}\}$. In this section, we also calculated the free energy for the **DBA-1**$^{-2}$+Gal reaction. Increasing the number of trials N reduces the statistical noise for sampling the free energies $G^\circ_{\text{CE}}(X)$.

Tables 4.4, 4.5, and 4.6 show the results of $\Delta_r G^\circ_{\text{CE}}$, IQR, and the signed error $\Delta\Delta_r G^\circ_{\text{CE-Exp}}$ with respect to the experimental reference (CD or PT). For the reactions **DBA-1**$^{-2}$+Glc and **DBA-2**+Glc, the statistical values fluctuate in an energy range less than 1 kcal/mol with the number of trials N, coinciding with a shape convergence to a monomodal distribution, as shown in the Figures A.15(a) and A.15(b). For the **DBA-1**$^{-2}$+Gal reaction, Figure A.15(c)

Table 4.4: Free energies for the reaction **DBA-1**$^{-2}$ + Glc $\overset{K}{\rightleftharpoons}$ **DBA-1.Glc**$^{-2}$ + 4 H_2O, calculated using the setting B for CE refinement, seven water molecules for microsolvation $(H_2O)_7$, and increasing number of trials N for the molecular free energies $G^{\circ}_{\mathrm{CE},i}(X)$. $\Delta_r G^{\circ}_{\mathrm{CE}}$, IQR, and $\Delta\Delta_r G^{\circ}_{\text{CE-CD}}$ reported in kcal/mol.

N	$\Delta_r G^{\circ}_{\mathrm{CE}}$	IQR	$\Delta\Delta_r G^{\circ}_{\text{CE-CD}}$
3	-2.5	1.4	3.3
5	-2.1	1.6	3.7
7	-2.6	1.9	3.2
10	-2.4	1.9	3.4

Table 4.5: Free energies for the reaction **DBA-2** + Glc $\overset{K}{\rightleftharpoons}$ **DBA-2.Glc** + 4 H_2O, calculated using the setting B for CE refinement, seven water molecules for microsolvation $(H_2O)_7$, and increasing number of trials N for the molecular free energies $G^{\circ}_{\mathrm{CE},i}(X)$. $\Delta_r G^{\circ}_{\mathrm{CE}}$, IQR, $\Delta\Delta_r G^{\circ}_{\text{CE-PT}}$, and $\Delta\Delta_r G^{\circ}_{\text{CE-NMR}}$ reported in kcal/mol.

N	$\Delta_r G^{\circ}_{\mathrm{CE}}$	IQR	$\Delta\Delta_r G^{\circ}_{\text{CE-PT}}$	$\Delta\Delta_r G^{\circ}_{\text{CE-NMR}}$
3	-6.6	2.3	-2.0	-1.1
5	-6.2	2.3	-1.6	-0.7
7	-6.4	2.0	-1.8	-0.9
10	-6.4	2.0	-1.8	-0.9

shows an initial bimodal shape for $\{\Delta_r G^{\circ}_{\mathrm{CE},x}\}$, converging to a monomodal distribution with increasing the samples N, and showing variations of less than 1 kcal/mol for the statistical values when $N \geq 5$. The initial bimodal distribution of $\{\Delta_r G^{\circ}_{\mathrm{CE},x}\}$ is due to outliers in the sampled $G^{\circ}_{\mathrm{CE},i}(X)$. Microsolvated configurations with outlier energy values have also been reported for the SN_2 reaction CH_3Cl + NH_3 in aqueous media, where some configurations may even eliminate the reaction energy barrier of 23.5 kcal/mol.[164] However, in our approach, the impact of energy outliers diminishes as the number of samples (N) increases, leading to an improvement in accuracy, as shown in Table 4.6.

Finally, for the set of three reactions in this section, we report for our workflow an average absolute error $\langle|\Delta\Delta_r G^{\circ}_{\text{CE-Exp}}|\rangle$ = 2.2 kcal/mol respect the experimental references, and average $\langle\text{IQR}\rangle$ = 2.4 kcal/mol, when using ten trials ($N = 10$) of molecular free energies $G^{\circ}_{\mathrm{CE}}(X)$, setting B, and seven water molecules for microsolvation of all the solutes. With the above-mentioned settings, the performance of our workflow is remarkable considering that, in general, experimental setups hardly achieve measurement precision of 1-2 kcal/mol.[90] Most importantly, the number of trials can significantly impact the accuracy and it is critical to ensure that the distribution of reaction-free energies is approximately monomodal, especially

Table 4.6: Free energies for the reaction **DBA-1**$^{-2}$ + Gal $\overset{K}{\rightleftharpoons}$ **DBA-1.Gal**$^{-2}$ + 4 H_2O, calculated using the setting B for CE refinement, seven water molecules for microsolvation $(H_2O)_7$, and increasing number of trials N for the molecular free energies $G^{\circ}_{\mathrm{CE},i}(X)$. $\Delta_r G^{\circ}_{\mathrm{CE}}$, IQR, and $\Delta\Delta_r G^{\circ}_{\mathrm{CE\text{-}CD}}$ reported in kcal/mol.

N	$\Delta_r G^{\circ}_{\mathrm{CE}}$	IQR	$\Delta\Delta_r G^{\circ}_{\mathrm{CE\text{-}CD}}$
3	-9.3	4.3	-5.1
5	-6.5	4.0	-2.3
7	-6.7	3.9	-2.5
10	-6.1	3.5	-1.9

if experimental data is unavailable.

4.2 Stable Products from the Reaction of Sucrose with aBA Molecules

We used the sucrose conformer from its crystalline structure, shown in Figure 1.1, to identify the hydroxyl groups in sucrose which form stable products with PBA and BOB molecules. Computational studies[35,109] indicate that the most stable conformer of sucrose in water closely resembles its crystalline form,[34] dominating the dynamics in aqueous media, despite the presence of three additional conformers.[109]

In Figure 3.4, we assumed a tetrahedral geometry for the boron atoms in both the reactants and products of aBAs reactions with sucrose. As a result, the boron atoms in the products became new chiral centers. We also screened the different chiral configurations of the boron atom formed after the reaction, as it is well known that chirality influences the reactivity of the system. Some of the product structures, along with their corresponding chiralities, are shown in Figure 4.4.

The reaction enthalpies ($\Delta_r H^{\circ}$), calculated from equation (3.3), are reported in Tables 4.7 and 4.8 for PBA and BOB, respectively. None of the reactions of PBA or BOB with hydroxyl groups at the 2,3-positions resulted in negative values for $\Delta_r H^{\circ}$, indicating that these products are less stable than the reactants. All 1,2-hydroxyl groups in sucrose have a trans configuration, which prevents the formation of a stable envelope conformation for a five-membered ring with sp^3 hybridized atoms.[165] Based on these results, we did not continue screening reactions involving aBAs with the 1,2-hydroxyl groups.

For reactions where PBA binds to sucrose at the 4,6- and 1',3'-hydroxyl groups, the

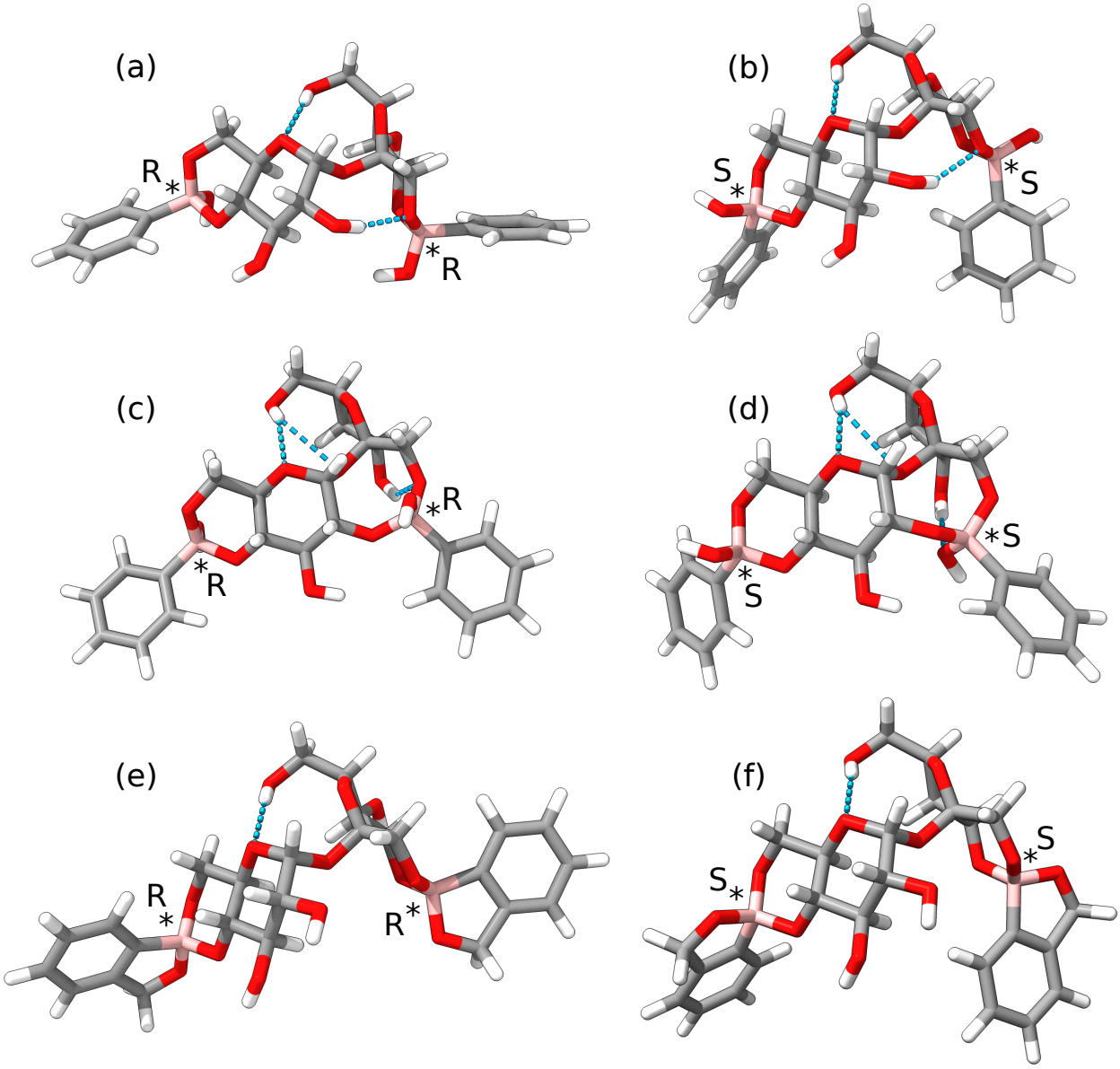


Figure 4.4: Representative product structures from the reaction of sucrose with PBA and BOB. The letters "R" and "S" indicate the chirality of the boron atom. Dashed blue lines mark the hydrogen bond interactions. Subfigures (a) and (b) show the optimized structures of two PBA molecules bound to sucrose at the positions 4,6- on the glucose unit and the positions 1',3'- on the fructose unit. Subfigures (c) and (d) depict one PBA molecule bound at the 4,6-positions of the glucose unit and the other bound at the position 2- of glucose and the position 3'- of fructose. Subfigures (e) and (f) correspond to subfigures (a) and (b), respectively, but with PBA replaced by BOB.

Table 4.7: Reaction enthalpy changes $\Delta_r H°$ (kcal/mol) for the PBA bonded at different hydroxyl groups in sucrose. Letters R and S are used to label the chirality of the boron atom in the borodiester ring with sucrose, as shown in Figure 4.4. The blue and red letters are the labels for the carbon atoms in the glucose and the fructose unit, respectively, as used in the Figure 1.1. Calculations at the PBEh-3c level of theory with the SMD implicit solvent model for water.

		2,1'-		1',3'-		4',6'-	
		R	S	R	S	R	S
2,3-	R			3.4	3.1		
	S			4.1	5.1		
4,6-	R	3.9	-6.8	-5.1	-9.8	-1.8	-4.3
	S	-2.3	-3.0	-7.9	-6.9	-1.7	0.8

calculated $\Delta_r H°$ values are negative, indicating that the borodiester compounds are more stable than the reactants. However, binding of BOB at the same positions did not yield stable products. The hydroxyl groups at the 4,6- and 1',3'-positions have a trans 1,3-configuration, which allows for the formation of a six-membered ring in a stable chair conformation with sp^3 hybridization when bound to PBA.[36] This may explain the negative $\Delta_r H°$ values observed for PBA+Sucrose reaction in Table 4.7. In contrast, BOB likely lacks sufficient rotation of its phenyl ring to reduce respulsive non-covalent interactions with sucrose atoms, which may account for the instability of the BOB-sucrose products compared to PBA-sucrose case.

Figures 4.4(a) and (b) depict two representative structures for the PBA-sucrose products, while Figures 4.4(e) and (f) show the BOB-sucrose products. In both cases, the six-membered rings adopt the chair conformation, but a key difference is the greater rotational flexibility of the phenyl ring in PBA, particularly in the R-chiral configuration, which may contribute to the increased stability of PBA products compared to BOB.

The reaction of aBAs with hydroxyl groups at the 4',6'-positions also forms a six-membered ring in a chair conformation. Two representative structures for the reaction of PBAs with sucrose at the 4',6'- and 4,6-positions are shown in Figure 4.5, and the reaction enthalpies ($\Delta_r H°$) are reported in Table 4.7. Comparatively, the 4,6-/4',6'- products are less stable than the 4,6-/1',3'- products, despite all the six-membered rings adopting a chair conformation. Rozada and co-workers[35] reported that the stability of sucrose's most favorable conformation (as shown in Figure 1.1, Left) is due to intramolecular hydrogen bonds and stabilizing molecular orbital hyperconjugation across the glycosidic bonds. Some of these stabilizing interactions are absent on the optimized structures of the 4,6-/4',6'- products. In contrast, for the 4,6-/1',3'- products, most of these interactions are preserved, as evidenced by the

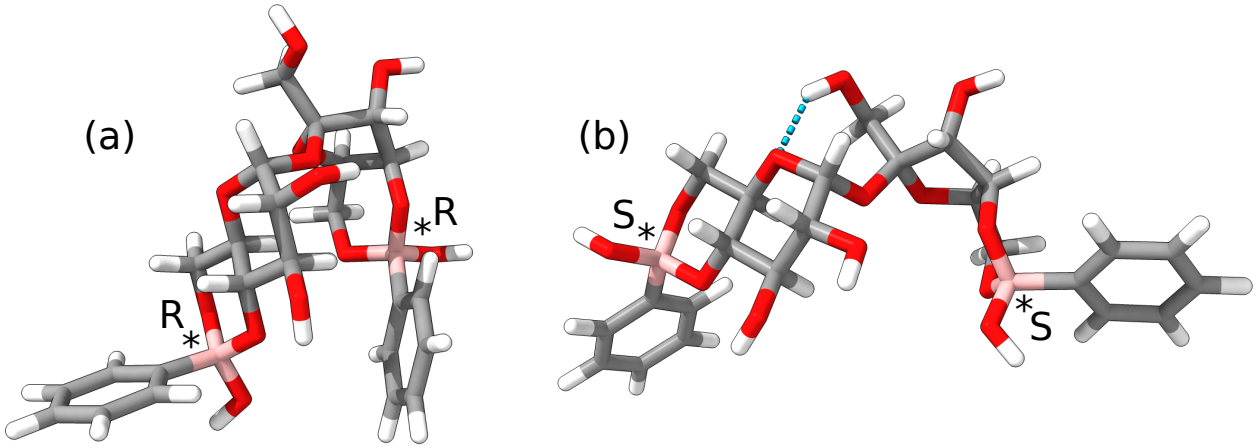


Figure 4.5: Representative structures for sucrose products with PBA at 4,6- and 4',6'-positions. The letters "R" and "S" indicate the chirality of the boron atom. Dashed blue lines mark the hydrogen bond interactions.

hydrogen bond interactions shown in Figures 1.1(Left) and 4.4. Based on these results, we did not perform analogous calculations for the BOB case.

Another potential reaction site for borondiester products with sucrose involves the 2-position on the glucose unit and the 1'-position on the fructose unit. The reaction enthalpies ($\Delta_r H°$) for PBA at these positions are generally negative, indicating favorable reactions in most cases. Figures 4.4 (c) and (d) depict two representative structures of PBA bound at the 2- and 1'-positions, with another PBA molecule bonded at the 4,6-positions. Despite the negative reaction enthalpies involving the 2- and 1'-positions, the formation of an eight-membered ring, which includes the highly flexible glycosidic bond, likely makes these products unstable due to the significant reduction of the conformational entropy.

Experimental evidence suggests that boric acid ($B(OH)_3$) reacts with sucrose at the 4,6-positions in aqueous media at pH 7, 25 °C, and a 1:1 molar ratio of boric acid to sucrose, forming a boronate compound, where boron adopts sp^3 hybridization with a negative formal charge.[166] The authors reported an overall equilibrium constant of $K_{eq} = 0.6$, corresponding to $\Delta_r G° = 1.3\,\text{kcal/mol}$ under these reaction conditions. However, given the acidity constant of boric acid ($pK_a = 9.14$),[166] these reaction conditions are not optimal for the reaction between $B(OH)_3$ and sucrose ($pH \geq pK_a$, see Figure 1.3).[47] Despite the positive value of $\Delta_r G°$ for the $B(OH)_3$ + sucrose reaction, the experimental findings of van der Berg and co-workers[166] demonstrate that sucrose can still exhibit reactivity with boronic acids in aqueous conditions, even under non-optimal pH conditions.

Table 4.8: Similar as reported in Table 4.7 for $\Delta_r H^\circ$ (kcal/mol), but for reactions of BOB with sucrose.

		2,1'-		1',3'-	
		R	S	R	S
2,3-	R			7.4	9.7
	S			6.8	7.5
4,6-	R	-4.1	-4.9	-1.0	0.2
	S	-1.4	-3.1	1.6	1.4

Finally, the optimized structures in Figure 4.4 can serve as reliable scaffolds for inverse design of DBA tweezers tailored for the selective detection of sucrose. The glycosidic bond in sucrose prevents mutarotation, meaning that isomerization after reaction with aBA is not expected, unlike in the cases of glucose and galactose (see Section 3.1.3).

4.3 *De Novo* Generation of Tweezers-Like Molecules for Binding to Sucrose

The optimized structures with negative values for the $\Delta_r H^\circ$ calculations (Tables 4.7 and 4.8) were used as templates for designing sucrose DBA tweezers. In this section, carbon-based molecular fragments are selected to complete the DBA tweezer and fix the aBA moieties at appropriate distance and orientation for reaction with sucrose. To achieve this, we scanned the dihedral angles related to the glycosidic bond and the phenyl rings within the PBA fragment. We then identified the molecular fragment that met the tolerance criteria necessary for closing the DBA tweezer, as outlined in Section 3.4.

In the case of BOB, one of the B−O−C covalent bonds fixes the phenyl rings, preventing their rotation. Consequently, no DBA tweezers based on BOB were generated using HostDesigner, as no linker fragments in its library met the tolerance criteria for matching with the attachment vectors of the BOB moieties. This emphasizes the need for some flexibility in the scaffold during the design process, even though high rigidity is likely a desired property for the designed DBA tweezers. As a result, we did not continue with the design of sucrose DBA tweezers based on BOB.

Table 4.9 presents the number of different DBA tweezers based on PBA. The scaffold with both S-chiral boron atoms, shown in Figure 4.4(b), produced the 1361 DBA tweezers. This is likely because the distance between the phenyl rings in this configuration is the

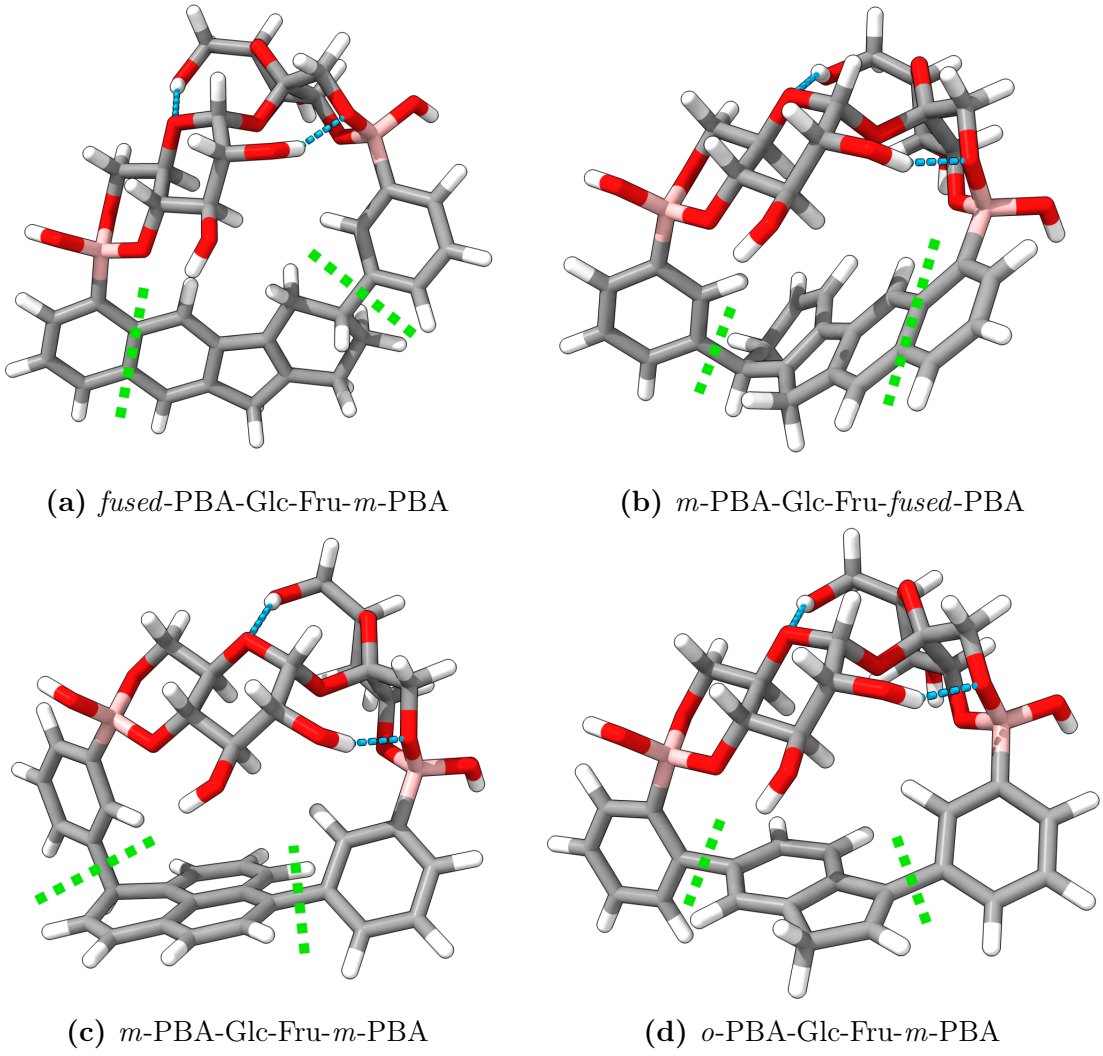
Table 4.9: Number of ligands generated for PBA based on the chirality of boron atoms, designed for specific binding to sucrose. The averaged from Tanimoto, Dice, and Cosine pairwise similarities scores,[137] using ECFP6 fingerprints, are reported in parentheses. Values closer to 1, indicate more similarity between the ligands.

		2,1'-		1',3'-	
		R	S	R	S
2,3-	R			–	–
	S			–	–
4,6-	R	–	10 (0.46)	0	0
	S	2 (0.69)	2 (0.31)	0	1361(0.28)

shortest among the different chiral configurations of the boron atoms. This allows fragments of various sizes to meet the matching criteria to close the DBA tweezer at different dihedral angles of the phenyl rings and glycosidic bond. The values in parentheses in Table 4.9 assess the topological similarity between the DBA tweezers. The DBA tweezers generated from the scaffold in Figure 4.4(b) also exhibit high diversity, as indicated by a low similarity score of 0.28. This is an important feature of the design process with the evolutionary algorithms. A highly diverse initial population of DBA tweezers enables a broader exploration of chemical space, increasing the likelihood of discovering an optimal tweezer.

The 1361 tweezers produced with the scaffold in Figure 4.4(b) can be classified into four types, as shown in Figure 4.6. They differ in how the PBA moieties are bonded to the linker. We used the prefixes *ortho-* (*o-*) and *meta-* (*m-*) to indicate that the boronic group is at the position 1,2-, or 1,3- relative to the bonding position of the linker on the aromatic ring of the PBA moiety. The term *fused-* means the PBA's phenyl ring is fused with the linker. Additionally, since the relative distance and orientation of the boronic groups are important for selectivity toward sucrose, we also distinguished to which side of sucrose the PBA moiety is bonded. For example, in Figure 4.6(a), the caption *fused*-PBA-Glc-Fru-*m*-PBA means that the phenyl ring of the PBA bonded to the Glucose (Glc) side of sucrose is fused with the linker, while the linker is bonded in a 1,3-position relative to the boronic group bonded to the Fructose (Fru) side of sucrose.

The type of DBA tweezers produced with HostDesigner depends on the library of fragments used.[60,61] Generative models for linkers[167] may help to overcome the limitations of a finite fragment database. These models can enhance linker diversity and enable the design of DBA tweezers for boron centers with chiralities beyond the S,S-configuration. However, when the distances between the phenyl rings increase, linker fragments with more atoms are



(a) *fused*-PBA-Glc-Fru-*m*-PBA

(b) *m*-PBA-Glc-Fru-*fused*-PBA

(c) *m*-PBA-Glc-Fru-*m*-PBA

(d) *o*-PBA-Glc-Fru-*m*-PBA

Figure 4.6: Representative structures of tweezer DBA-sucrose complex produced with Host-Designer. The dashed green lines serve as visual guides to indicate the separation between the PBA moieties and the linker region.

required, which can complicate the synthesis process.[139,168] Additionally, generative models may encounter difficulties in producing compounds that are synthetically accessible.[66]

The thermodynamic quantities reported, such as the reaction enthalpy ($\Delta_r H^\circ$) and reaction free energy ($\Delta_r G^\circ$), provide insights into the stability of the reaction products and the spontaneity of the reaction at a given temperature. To continue the analysis of the reactivity of the DBA tweezers with sucrose, the next section presents our results about the barriers and mechanism for the reaction of sucrose with PBA.

4.4 Reaction Mechanisms of Sucrose with PBA

We limited the analysis of the reaction mechanism of PBA with sucrose on the 1',3'- and 4,6-hydroxyl groups (see Figure 1.1), and the S-chirality of the boron atom (Figure 4.4(b)), as these configurations yielded the majority of DBA tweezers for sucrose, as shown in Table 4.9.

From exploratory calculations of the PES for the sucrose+PBA system using the NEB-CI method, we propose that the reaction begins with the deprotonations of the primary hydroxyls 6- of glucose moiety and 1'- of fructose moiety for reaction conditions studied. This step leads to the formation of the first B−O−C bond, as shown in Figures 4.7 and 4.8 for reaction condition pH $\geq$ p$K_{a,\mathrm{PBA}}$, and Figures C.1 and C.2 for reaction condition pH $<$ p$K_{a,\mathrm{PBA}}$.

For the reaction mechanism of the sucrose + PBA system presented in this work, all the TSs involve the deprotonation of hydroxyl groups in sucrose to produce H_2O and the B−O−C bond. Therefore, it is pertinent to begin the discussion with the reported experimental findings on the acidity of sucrose's hydroxyl groups.[169,170] In high-water-content solutions, the acidity of the −OH groups in sucrose is approximately equal,[169] with high rate coefficients from 10 to 20 s^{-1} for proton exchange between the sucrose −OH groups and solvent molecules in a 50/50 H_2O/D_2O solution.[170] However, the rate exchanges for the hydrogen atoms between H_2O and the primary −OH sucrose groups are notably faster in solutions of 3:1(v/v) acetone-d_6/water, as calculated from experiments of NMR with total correlation spectroscopy.[169] These faster exchanges result in broadened peaks broaden peaks in the NMR spectra for 2:1(v/v) acetone-d_6/water solutions.[169] The primary −OH are more exposed to solvent molecules, facilitating rapid proton exchange and broadening their corresponding NMR peaks.[169] Consequently, initiating the reaction at the primary −OH groups at positions 6- and 1-' of sucrose with PBA is a reasonable and supported approximation.

Intramolecular hydrogen bonding is another factor that may potentially hinder the re-

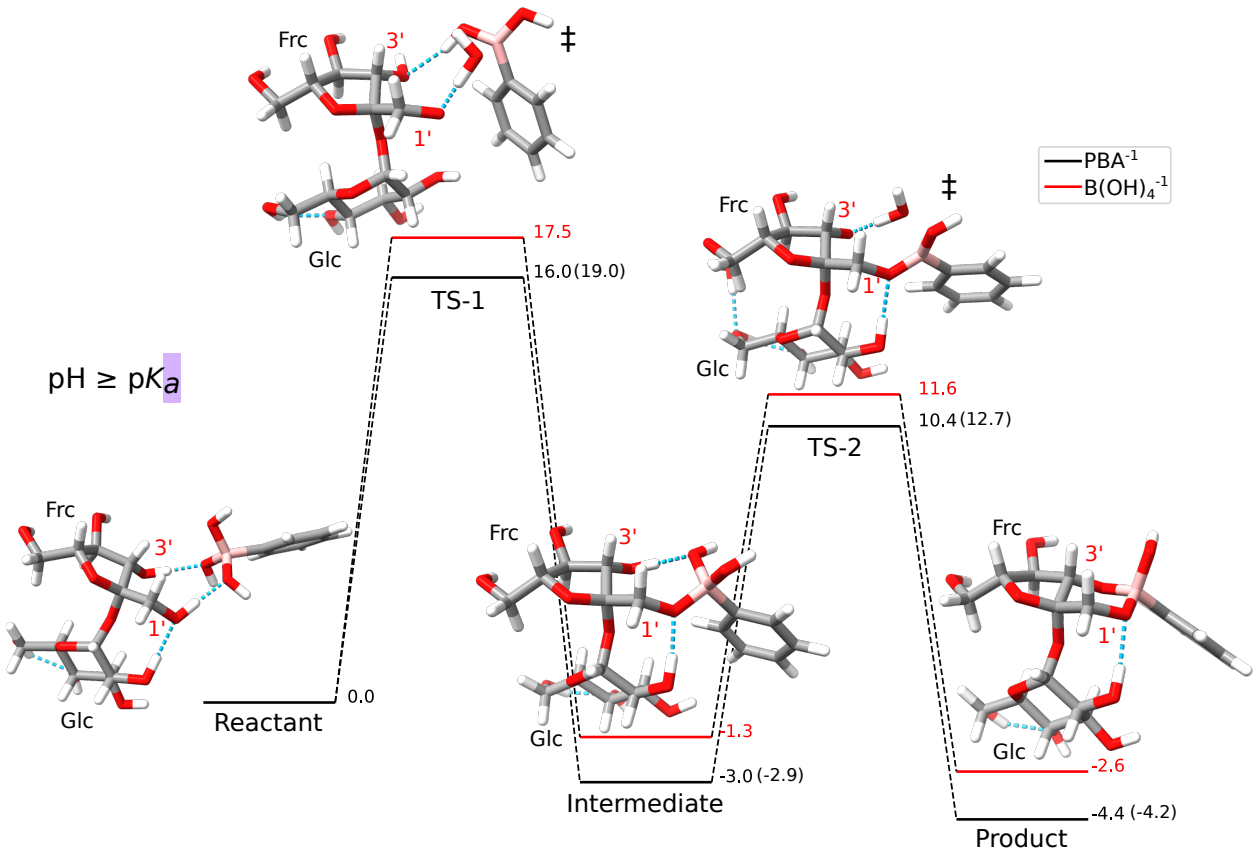


Figure 4.7: Reaction barriers ($\Delta G^{\ddagger,\circ}$) and reaction-free energies ($\Delta_r G^\circ$) for the reaction of PBA^{-1} (black line) and $B(OH)_4^-$ (red line) with the fructose moiety of sucrose at reaction conditions pH $\geq$ pK_a of the boronic compound. Calculations were carried out at the r^{2}SCAN-3c level of theory, with values in parentheses indicating results from ωB97M-D4/def2-QZVPP//r^{2}SCAN-3c calculations for Self-Interaction Error (SIE) corrrections, and the SMD solvent model for water. All vibrational frequencies were obtained at the r^{2}SCAN-3c level. For PBA^{-1}, the imaginary frequencies for transition states TS-1 and TS-2 are 213.03i cm^{-1} and 225.16i cm^{-1}, respectively. For $B(OH)_4^-$, the corresponding imaginary frequencies are 303.95i cm^{-1} for TS-1 and 265.07i cm^{-1} for TS-2. Dashed blue lines highlight the hydrogen bonds.

action of sucrose with DBA tweezers. NMR ^{1}H J-coupling experiments and computational chemistry calculations[109,171,172] have demonstrated that intramolecular hydrogen bonds in sucrose are relatively weak in high-water-content solutions. Consequently, sucrose exhibits significant flexibility in aqueous environments, with its −OH groups (almost) freely rotating to adopt favorable orientations for reactions with pre-coordinated PBA moieties in DBA tweezer compounds.

At reaction condition pH $\geq$ p$K_{a,\mathrm{PBA}}$, the boronic group in the PBA^{-1} reacts with sucrose in a sp^3 tetrahedral (R−B(OH)$_3^-$) configuration (Figures 4.7 and 4.8). TS-1 and TS-2 involve the hydrogen transfer from the −OH groups in sucrose to a −OH in PBA to produce a water molecule and the formation of the B−O−C bond, being the first B−O−C bond formation the rate-determining step in all the reaction mechanisms. The hydrogen transfer from sucrose's −OH groups to the tetrahedral R−B(OH)$_3^-$ group, emulates an SN_1-type reaction with transitory formation of an alkoxy ($-O^-$) group at the TSs in Figures 4.7 and 4.8. This is confirmed by the Self-Interaction Error (SIE) correction performed to the reaction barriers ($\Delta_r G^{\circ,\ddagger}$) at the ωB97M-D4/def2-QZVPP//r^{2}SCAN-3c level of theory. The r^{2}SCAN-3c method is prone to over-delocalize the electrons in the $-O^-$ group due to an artificial electron self-interaction, over-stabilizing the TSs and reducing the reaction barriers $\Delta_r G^{\circ,\ddagger}$, but this errors are corrected with ωB97M-D4/def2-QZVPP single point calculations.[90]

Meanwhile, at reaction condition pH $<$ p$K_{a,\mathrm{PBA}}$, the boronic group reacts with sucrose in a sp^2 trigonal (R−B(OH)$_2$) configuration (Figures C.1 and C.2). The hydrogen transfer from sucrose's −OH groups to trigonal R−B(OH)$_2$ group in a concerted fashion, similar to an SN_2-type reaction. Consequently, SIE corrections to $\Delta_r G^{\circ,\ddagger}$ are not important in these cases (see Figures C.1 and C.2).

The hydrogen transfer observed in all reported TSs in this work may explain the experimental evidence for the lower reactivity of deprotonated diols compared to their protonated counterparts in forming boron ester compounds.[52,173] Hydrogen transfer between −OH groups is the initial step in the condensation reaction pathway, which involves H_2O elimination followed by B−O−C bond formation. Consequently, deprotonated $-O^-$ groups in diols are unable to react via the condensation pathway and encounter a higher reaction barrier when forming the B−O−C bond.

Boronic groups exhibit higher reactivity with sucrose in the tetrahedral configuration (R−B(OH)$_3^-$) compared to the trigonal configuration (R−B(OH)$_2$). The reaction barriers $\Delta_r G^{\circ,\ddagger}$ are lower than 20 kcal/mol in the tetrahedral configuration. In contrast, reactions in the trigonal configuration involve reaction barriers $\Delta_r G^{\circ,\ddagger}$ higher than 25 kcal/mol. This coin-

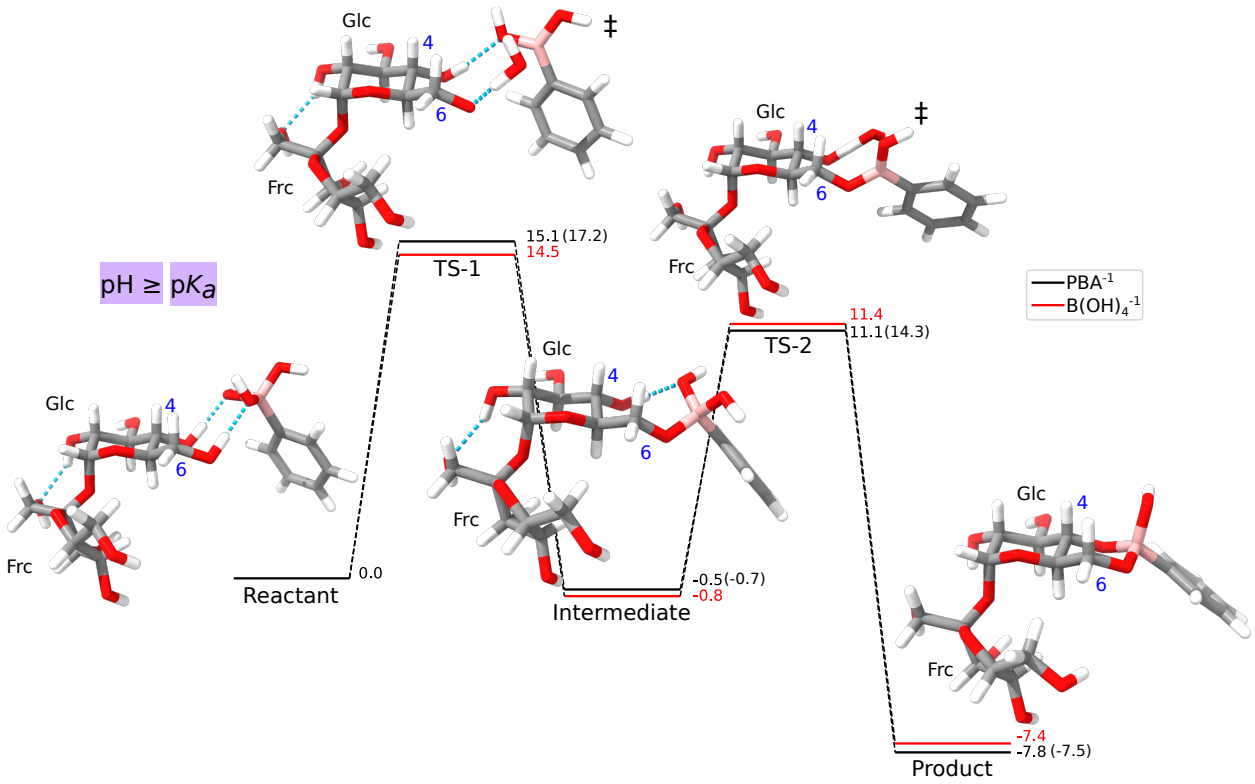


Figure 4.8: Same as Figure 4.7 but for PBA^{-1} (black line) and $B(OH)_4{}^{-1}$ (red line) reacting with the glucose moiety of sucrose at reaction conditions pH $\geq$ pK_a of the boronic compound. For PBA^{-1}, the imaginary frequencies for transition states TS-1 and TS-2 are 305.88i cm^{-1} and 455.37i cm^{-1}, respectively. For $B(OH)_4{}^{-1}$, the corresponding imaginary frequencies are 259.19i cm^{-1} for TS-1 and 308.14i cm^{-1} for TS-2. Dashed blue lines highlight the hydrogen bonds.

cides with experimental reports where the boronate ion $R{-}B(OH)_3{}^-$ dominates the reaction kinetics with diols at reaction conditions of $pH > pK_{a.aBA}$ of aBA compounds.[19,52,173–177]

Table 4.10 present the canonical rate coefficients ($k(T)$) and half-life times ($t_{1/2}$) for the rate-determining steps for mechanism at reaction conditions pH $\geq$ p$K_{a,PBA}$, presented in Figures 4.7 and 4.8. The analogous values for reaction conditions pH $<$ p$K_{a,PBA}$ are presented in Table C.1. The $k(T)$ values for the reaction of the PBA^{-1} with sucrose are comparable with those reported for aBA compounds reacting with fructose,[52] ethylene glycol,[174] and propylene glycol[174] under reaction conditions where pH $>$ p$K_{a,aBA}$. The higher reactivities and more stable product of PBA^{-1} and $B(OH)_4{}^-$ reacting with the glucose side of sucrose, also coincide with the reaction product proposed from NMR experiments[166] for the reaction $B(OH)_3$ + sucrose at pH = 7.0 and 25°C. The absence of experimental reports on boron

Table 4.10: Reaction rate coefficients ($k(T)$)[a] and half-life times ($t_{1/2}$)[b] for the reactions of PBA with sucrose under conditions of pH $\geq$ p$K_{a,\mathrm{PBA}}$. $\Delta_r G^{\circ,\ddagger}$ is 17.2 kcal/mol[c] for the glucose-side reaction and 19.0 kcal/mol[c] for the fructose-side reaction.

	Glucose-side		Fructose-side	
$T(°C)$	$k(T)(s^{-1}M^{-1})$	$t_{1/2}\,(min)$	$k(T)(s^{-1}M^{-1})$	$t_{1/2}\,(min)$
25	1.6	7.2	7.6×10^{-2}	151.2
30	2.6	4.5	0.1	89.1
35	4.1	2.8	0.2	53.4
40	6.4	1.8	0.4	32.5
50	15.0	0.8	0.9	12.7

[a] $k(T)$ calculated with the Eyring equation.[178,179] [b] $t_{1/2} = \ln(2)/k(T)[\mathrm{PBA}]_0$, where $[\mathrm{PBA}]_0 = 1.0 \times 10^{-3}\,\mathrm{M} = [\mathrm{Suc}]_0$.[180] [c] $\Delta_r G^{\circ,\ddagger}$ values calculated at the ωB97M-D4/def2-QZVPP//r^{2}SCAN-3c level of theory.

ester compounds formed with the fructose side of sucrose (Figure 4.7) may be attributed to a higher reaction barrier and lower stability compared to those formed with the glucose side (Figure 4.8). However, increasing the temperature to approximately 40°C, and maintaining the condition pH $\geq$ pK_a, may be required to accelerate the reaction kinetics on the fructose side, enabling the formation of boron ester compounds within a timescale of minutes.

Reaction half-times on the scale of minutes are essential for achieving the temporal resolution required to correlate plant responses to biotic and abiotic stress with sucrose concentrations. However, to determine whether a reaction half-time of two hours is sufficiently rapid for continuous sucrose monitoring in plants, additional information is needed—specifically, the limits of detection and quantification, which heavily depend on the transduction signal used to detect the DBA tweezer-sucrose product.

Experimental evidence suggest that trigonal R$-$B(OH)$_2$ aBA compounds are reactive with diols even under conditions where pH > pK_a. However, these reactions are exoergic, and the reverse reaction, Boronester $\longrightarrow$ R$-$B(OH)$_2$ + diol, tends to dominate the kinetics.[52,174] Computational studies have shown that water molecules catalyze the reaction of trigonal boronic acids R$-$B(OH)$_2$ with diols.[181,182] Understanding the reactivity of R$-$B(OH)$_2$ with sucrose is crucial for sensor design, as changes in R$-$B(OH)$_2$ concentration due to its reaction with sucrose—resulting in the formation of negatively charged boron ester compounds—can serve as a transduction signal for sucrose quantification.[183] Nonetheless, a comprehensive kinetic model incorporating all acid/base equilibria of boronic acids, intermediates, and boron ester compounds is needed to fully elucidate the reaction kinetics of aBA with sucrose.

4.5 Sucrose DBA Tweezers Improvement with GB-GA Method

In Section 4.3, we presented our results for the inverse design of DBA tweezers for selective sucrose detection. However, further enhancements are necessary to ensure optimal performance for sucrose quantification under physiological conditions.

In this section, we report the results of the holistic improvement of the DBA tweezers for sucrose using the GB-GA method (Section 3.5) and the multi-objective optimization function (3.6).

4.5.1 GB-GA Parameter Optimization for Exploration and Exploitation of the Chemical Space

Accurate calculation of the reaction free energy $\Delta_r G^\circ$ for DBAtweezer + carbohydrate systems (as presented in Section 4.1) is computationally prohibitive for use in the GB-GA process due to its high computational cost. To address this limitation, we modified Setting B for CENSO sorting (see Table 3.2) to enhance the throughput of $\Delta_r G^\circ$ calculations during GB-GA optimization. Specifically: (i) $\Delta_r G^\circ$ values were estimated using the SMD implicit solvation model for CENSO sorting; (ii) conformational entropies were directly obtained from conformers generated by CREST, without extrapolating to achieve near-complete ensemble sampling;[184] and (iii) conformational ensembles were generated using a short metadynamics simulation to sample the conformational space quickly. These adjustments accelerate the screening process, trading some accuracy for efficiency in the GB-GA evolutionary workflow. Additionally, at this stage, we focus on relative differences between DBA tweezers rather than accurate absolute values of $\Delta_r G^\circ$.

To initiate the evolutionary process, hyperparameters such as mutation rate (MUT) and selection pressure (η) were tuned to balance the exploration of chemical space and exploitation of high-performing DBA tweezers. Other parameters, including population size and replacement factor, were set to 32 and 0.5, respectively, following the recommendations in the reference 185. Coefficients c_1, c_2, c_3, and c_4 in Equation 3.6 were assigned values of 0.50, 0.20, 0.15, and 0.15, respectively, to prioritize $\Delta_r G^\circ$, optimizing the product stability, and reaction spontaneity between DBA tweezers and sucrose; and $\log P_{\mathrm{o/w}}$ in second place. Convergence was defined as the point at which the best-performing tweezer remained unchanged for ten consecutive iterations. Due to computational limitations, some parameters

(e.g., parental selection methods) were not optimized.

Figure B.1 presents the initial population (Generation 0) of sucrose DBA tweezers, corresponding to a sample of the sucrose DBA tweezers designed in Section 4.3. Table B.4 shows the $\Delta_r G^\circ$, $\log P_{\mathrm{o/w}}$, maximum pK_a, and SA values for Generation 0. The initial population is composed of carbon, hydrogen, and boronic groups, exhibiting poor water solubility with $\log P_{\mathrm{o/w}} > 4.0$ and suboptimal maximum pK_a values of 8.8 for the reaction of sucrose with PBA compounds at pH = 7.4. Some tweezers showed positive $\Delta_r G^\circ$, indicating minimal or no product formation with sucrose, and being unsuitable for sucrose detection. Positive $\Delta_r G^\circ$ values confirm that the structure-based design of DBA tweezers is an incomplete methodology.

The results for the GB-GA evolutionary process with different mutation rates (MUT) and selection pressure parameters η, are shown in Figure 4.9(a). Although the parameter screening was performed with the same initial population, the variations in the tweezer properties shown in Figure 4.9 at Generation 0 are due to the stochastic nature of the conformational sampling with the CREST protocol. However, these variations are close to the limit of chemical accuracy for $\Delta_r G^\circ$ (around 2.5 kcal/mol) and less than one unit for $\log P_{\mathrm{o/w}}$.

Regardless of the parameter combination, the median score of the best-performing tweezers (Figure 4.9(a)) consistently decreases during the iterative process, aligning with the minimization of the objective function J (Equation (3.5)) for the DBA tweezer optimization. To identify the most cost-effective combination of parameters MUT and η for the GB-GA optimization, we evaluated the median number of tweezers in the top 100 elite across different settings.[185] The top 100 elite were selected from the final populations of all runs and parameter combinations. Percentage coverage was used to assess the algorithm's ability to explore diverse local minima while optimizing various molecules.[185] Figure 4.9(b) presents the percentage coverage for all tested MUT and η combinations. Based on the median score (Figure 4.9(a)) and percentage coverage (Figure 4.9(b)), the combination MUT = 0.5 and $\eta = 1.5$ offers the most balanced trade-off between chemical space exploration and optimization of high-performing DBA tweezers. High mutation rates (e.g., MUT = 0.8) tend to trap the GB-GA algorithm in local minima, leading to few optimal tweezers. Evaluating the influence of η independently from MUT is challenging; however, within the range $1.0 \leq \eta \leq 2.0$, $\eta = 1.5$ appears to achieve an optimal balance between exploration and exploitation.[140] This value ensures an effective selection bias of the parental DBAtweezers for mutation and crossover operations, promoting both diversity and the refinement of high-performing candidates.

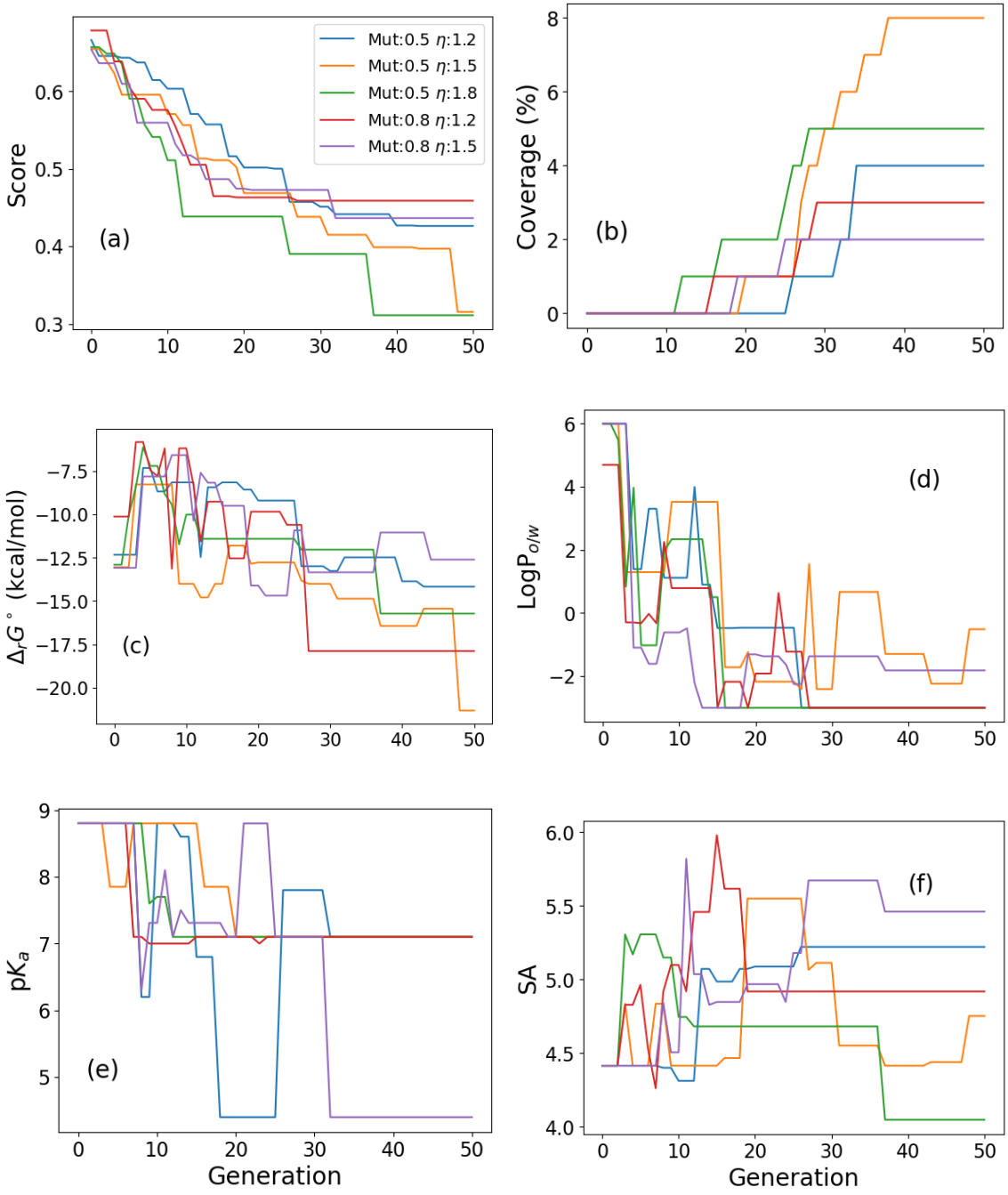


Figure 4.9: Median values of the top-performing sucrose DBA tweezers across generations with different ratios for mutation rates (MUT) and selection pressure (η) parameter in Equation (3.8). Each parameter combination for the evolutionary algorithm was run thrice with the same initial population of 32 sucrose DBA tweezers.

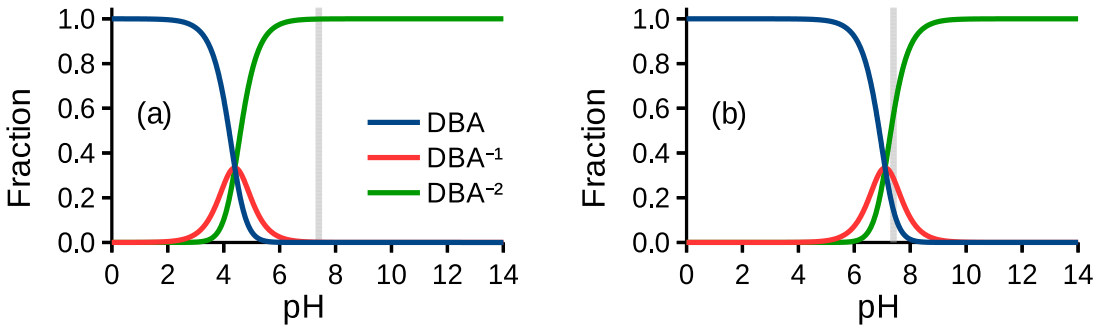


Figure 4.10: Speciation diagrams as a function of pH for a hypothetical DBA tweezer molecule with pK_a values of (a) 4.4 and (b) 7.1. In these diagrams, DBA represents the neutral species, DBA^{-1} denotes the molecule with one boronic group with a negative formal charge, and DBA^{-2} corresponds to the species with both boronic groups with negative formal charges. The dashed gray line indicates pH = 7.4 for reference.

The properties evolution of the best-performing sucrose DBA tweezers subject to optimization are shown in Figures 4.9(c)-(f). Although the $\Delta_r G^\circ$ (Figure 4.9(c)) and $\log P_{o/w}$ (Figure 4.9(d)) show a decreasing trend, occasional increases occur due the trade-off of conflicting properties, characteristic in multi-objective optimization problems.

The maximum pK_a values (Figure 4.9(e)) for the best performing, in all the cases, decrease from 8.8 to 7.1, which implies an improvement in the performance of the DBA molecules for reaction with sucrose at physiological conditions. Figures 4.10(a) and (b) show the speciation diagrams for two hypothetical DBA tweezers with pK_as equal to 4.4 and 7.1, respectively. For a DBA tweezer with maximum pK_as = 4.4, all the boronic groups are charged negatively ($R{-}B(OH)_3^{-1}$) at physiological pH. Meanwhile, for a DBA tweezer with maximum pK_as = 7.1, approximately 60% of the tweezer corresponds to the DBA^{-2} specie, and less than 15% corresponds to the neutral DBA specie. Therefore, even DBA tweezers with pK_as $\leq$ 7.1 may be useful for sucrose sensor because the majority of the tweezer DBA molecules may form stable compounds with the sucrose, following the reaction mechanism for the boron atoms in sp^3 hybridization, as shown in Figures 4.7 and 4.8 (Section 4.4).

Regarding the SA score, Figure 4.9(f) shows an increase of the SA score in most of the cases, beginning from SA = 4.5 and increasing around to 5.0, in the range of synthesis complexity of bio-active molecules or natural compounds.[139] An increasing value in SA, respect Generation 0, is related to the addition of heteroatoms and functional groups to the DBA tweezers that improve the solubility, reduce the maximum pK_a values, or even improve the non-covalent interaction between the DBA tweezer and sucrose, increasing at the same

time the synthesis complexity of the tweezers. The modification of the weights c_i in Equation 3.6, or the use of the Gaussian scalarization method,[66] may help to control the incremental synthesis complexity of the DBA tweezers during the GB-GA optimization.

The GB-GA methodology effectively optimizes DBA tweezers for sucrose sensing by addressing both solubility and reactivity. For example, by Generations 25 and 50 in Table B.4, all tweezers achieved negative $\Delta_r G^\circ$, indicating thermodynamically favorable interactions with sucrose. Additionally, improvements in $\log P_{\mathrm{o/w}}$ and maximum pK_a values ensure their optimal performance at physiological conditions. A threshold of $\log P_{\mathrm{o/w}} \leq 3.0$ is commonly in *in silico* drug design for molecules solubility and tissue absorption.[186] At the 25th iteration of the GB-GA optimization in Table B.4, most tweezers met this criterion. Similarly, maximum pK_a values below 7.4 make tweezers more suitable for sucrose sensing at physiological pH.

4.5.2 High Accuracy Calculation of Reaction-Free Energies for Sucrose DBA Tweezers

In this section, we complete all the steps of the workflow for the computational aided design of sucrose DBA tweezers, presented in Figure 3.11, by the high-accurate calculation of the reaction-free energies ($\Delta_r G^\circ$) for the best-performing tweezers from the GB-GA optimization described in the previous section.

To illustrate the advantages of our workflow, a subset of twelve tweezer, from the top 100 best-performin from the previous section, was selected for accurate calculation of $\Delta_r G^\circ$ using Setting B from Table 3.2. These twelve DBA tweezer molecules were filtered and selected based on the following criteria: $\log P_{\mathrm{o/w}} < 1.0$, maximum p$K_a < 7.4$, and SA < 6.0. Tweezers with chiral centers were excluded to avoid challenges associated with regioselectivity during synthesis and the complexities of purifying racemic mixtures. The properties of this final group of tweezers are detailed in Table 4.11, with their molecular structures shown in Figure 4.11.

Based on the results of Table 4.11, tweezers **1**-**4** and **8**-**10** react spontaneously with sucrose in aqueous solution at pH = 7.4, with $\Delta_r G^\circ_{\mathrm{CE}}$ values similar to those reported in section 4.1.5 for the reactions **DBA-1**$^{-2}$ + Glc, **DBA-2** + Glc, and **DBA-1**$^{-2}$ + Gal. Despite sucrose's conformational entropy ($S_{\mathrm{conf}} = 11.6\,\mathrm{cal/mol/K}$) is nearly twice higher compared to glucose ($S_{\mathrm{conf}} = 6.9\,\mathrm{cal/mol/K}$), the conformational entropy changes ($T\Delta S_{\mathrm{conf}}$) for sucrose tweezers reactions are comparable to those for glucose and galactose reactions in Table 4.1. This

Table 4.11: Properties of sucrose-DBA tweezers after optimization with the GB-GA algorithm. The reaction-free energies, $\Delta_r G^\circ_{\mathrm{CE}}$[a], were calculated using Setting B (see Table 3.2), with the values obtained during the GB-GA optimization shown in parentheses. $\Delta_r G^\circ_{\mathrm{CE}}$, IQR, $\Delta\overline{G}^\circ$, and $T\Delta S_{\mathrm{conf}}$ are reported in kcal/mol.

Tweezer	$\Delta_r G^\circ_{\mathrm{CE}}$[a]	$\log P_{\mathrm{o/w}}$	Max pK_a	SA	IQR	$\Delta\overline{G}^\circ$	$T\Delta S_{\mathrm{conf}}$	Score
1	-3.2 (-8.1)	-1.7	4.4	4.9	3.5	-6.4	-3.4	0.5611
2	-3.7 (-7.6)	-2.1	6.2	5.2	3.7	-6.5	-2.7	0.5942
3	-5.0 (-16.4)	-1.3	7.1	4.3	2.3	-7.8	-3.1	0.5985
4	-2.4(-6.7)	-1.7	6.1	4.9	2.4	-6.5	-4.4	0.6166
5	-0.9 (-8.9)	-2.3	6.2	5.3	2.7	-3.4	-2.7	0.6366
6	4.5 (-4.3)	-3.0[b]	7.0	3.6	1.6	-1.4	-6.1	0.6433
7	1.9 (-8.9)	-3.0[b]	6.2	5.1	3.9	-1.2	-3.2	0.6483
8	-0.9(-6.9)	-2.0	7.1	4.6	2.6	-4.7	-3.6	0.6541
9	-5.1 (-12.9)	0.9	7.0	5.0	1.6	-9.5	-4.4	0.6549
10	-3.0 (-11.8)	0.2	7.0	5.2	3.5	-6.5	-3.8	0.6766
11	0.6 (-7.8)	-1.5	7.1	4.9	4.5	-2.7	-3.4	0.6944
12	8.7 (-8.9)	-1.5	7.0	4.8	3.2	4.9	-4.0	0.6967

[a]The $\Delta_r G^\circ_{\mathrm{CE}}$ values were calculated with three trials samples of molecular free energies $G^\circ_{\mathrm{CE,i}}(X)$ for the reactants and products, as shown in section 3.1.2. [b]Values out of the range of the MinMax scaling in Equation (3.7).

highlights the effectiveness of including conformational entropy changes into the GB-GA optimization process to exclude flexible tweezers with high unfavorable entropy penalties ($T\Delta S_{\mathrm{conf}}$) that cannot be compensated with covalent and non-covalent interactions between the tweezer and sucrose ($\Delta\overline{G}^\circ$), as shown in Equation (1.30) and the second to last column in Table 4.11.

For tweezer reported in Table 4.11, the median difference between the reaction-free energies calculated in the GB-GA process and Setting B is 8.3 kcal/mol. This difference is in the same order of magnitude to the estimated differences of ∼5.0 kcal/mol[108] between microsolvation and the implicit solvent model SMD for $\Delta_r G^\circ$ in the reactions **DBA-1**$^{-2}$ + Glc, **DBA-2**+Glc, and **DBA-1**$^{-2}$+Gal, as reported in section 4.1.4. The large difference between the $\Delta_r G^\circ$s in tweezer **12** (17.7 kcal/mol) arises from intramolecular reaction involving deprotonation of the carboxylic group, reacting with the boronate group (R−B(OH)$_3^-$), forming R−B(OH)$_2$ and H_2O.

The modifications to setting B, described in the first paragraph of the previous section, increased the throughput for calculation of $\Delta_r G^\circ$ by a factor of ∼5. Despite the loss of accuracy of 8.3 kcal/mol, the order of the best-performing tweezers from the GB-GA cal-

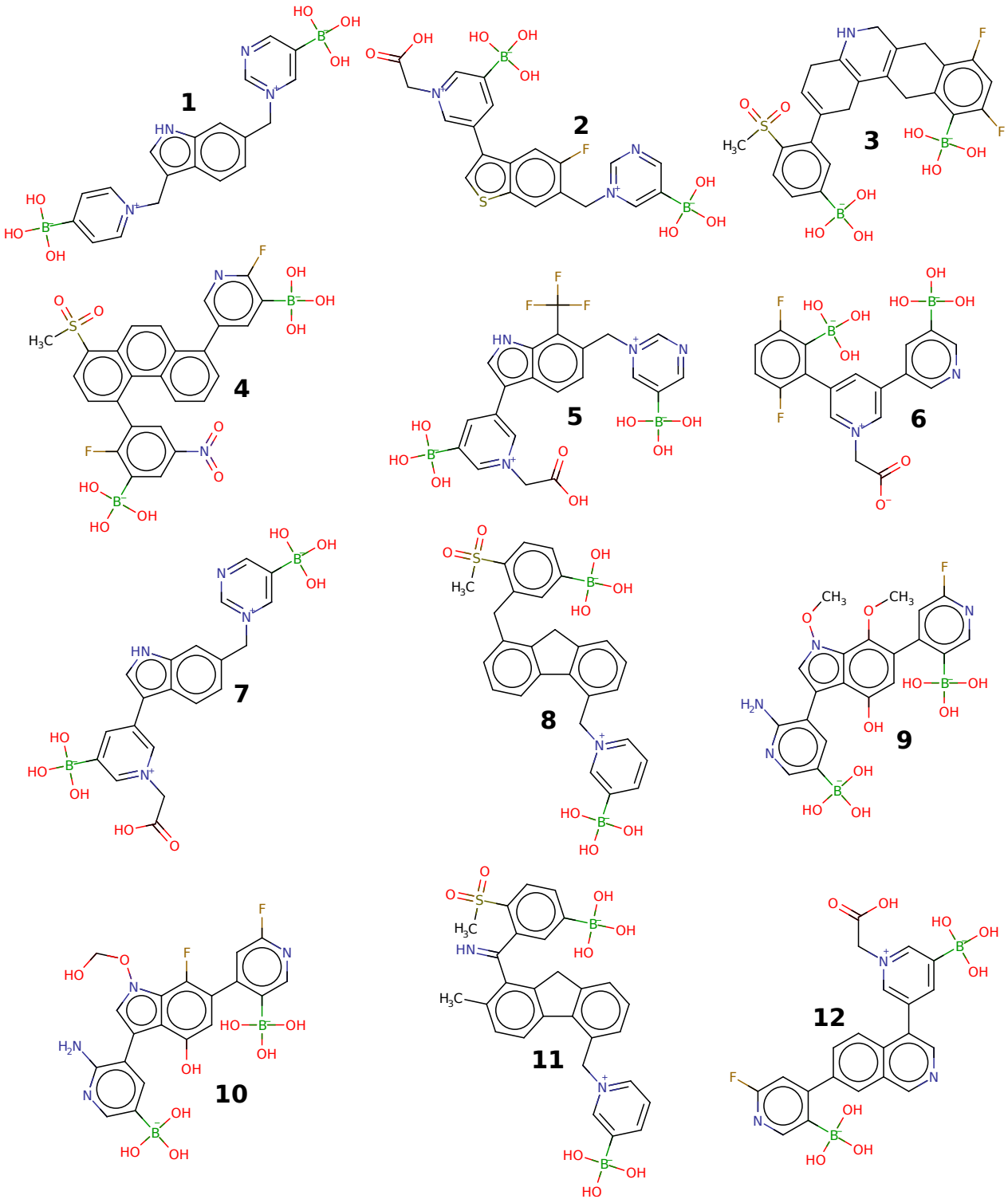


Figure 4.11: Selected group of DBA tweezer, optimized with the GB-GA algorithm, for detection of sucrose in physiological solutions.

culation, show modest changes after calculation of high accurate $\Delta_r G^\circ$, with a Spearman's rank correlation of 0.56 (excluding anomalous values for tweezer **12**). Therefore, the $\Delta_r G^\circ$ values obtained during GB-GA optimization provide a computationally efficient approach for estimating reaction-free energies, well-suited for the high-throughput demands of the GB-GA optimization.

It is noteworthy that, despite the rankings of tweezer **9** and **10**, their $\Delta_r G^\circ_{\mathrm{CE}}$ values indicate a spontaneous reaction with sucrose, while their $\log P_{\mathrm{o/w}}$ and $\mathrm{p}K_a$ values suggest good solubility and reactivity in aqueous media at pH 7.4. Conversely, the $\Delta_r G^\circ_{\mathrm{CE}}$ values for tweezers **5** to **8** imply that these tweezers are likely unsuitable for sucrose quantification, despite their higher rankings. This highlights the need for further refinement in selecting the weights c_i in Equation 3.6 to achieve a better balance among the tweezers' properties during the optimization and screening process.

The GB-GA optimization has enabled the design of DBA tweezers for selective sucrose detection in aqueous solutions. It also offers a framework for selecting tweezers with consideration for synthetic accessibility, thereby minimizing the challenges of experimental implementation of these DBA tweezers.

4.5.3 Selectivity of the DBA Tweezer for Sucrose

The core idea for tweezer design relies on fine-tuning the boronic group distances and relative orientations to react with a targeted carbohydrate. Comparing the boron-boron distances and dihedral angles of the boronic groups in both, the free tweezers and their corresponding cyclic boron ester complexes, can provide insights into the selectivity of the tweezers. Figure 4.12 presents these parameters for conformer ensembles of the free tweezers and the corresponding complexes formed with carbohydrates sucrose, glucose, and galactose.

From Figure 4.12, the boron atoms in the complexes formed with monosaccharides are separated by approximately 6.0 Å, while in the sucrose complexes, the distance between the boron atoms is approximately 7.9 Å, highlighting the need for a specific distance between the boron atoms to enable selective reactions with a target carbohydrate. The aBA groups in the non-complexed tweezers are free to rotate, which explains why the boron atom distances in the free tweezers show more variance compared to those in the complexes. Similarly to the B-B distances, the B-C-C-B dihedral angles in the complexes exhibit different values for optimal binding with monosaccharides or sucrose: dihedrals around 60° and 45° for glucose and galactose, respectively, and an optimal dihedral of -57° for sucrose.

None of the conformers of **DBA-1**$^{-1}$ or **DBA-2** align with the configurations required for

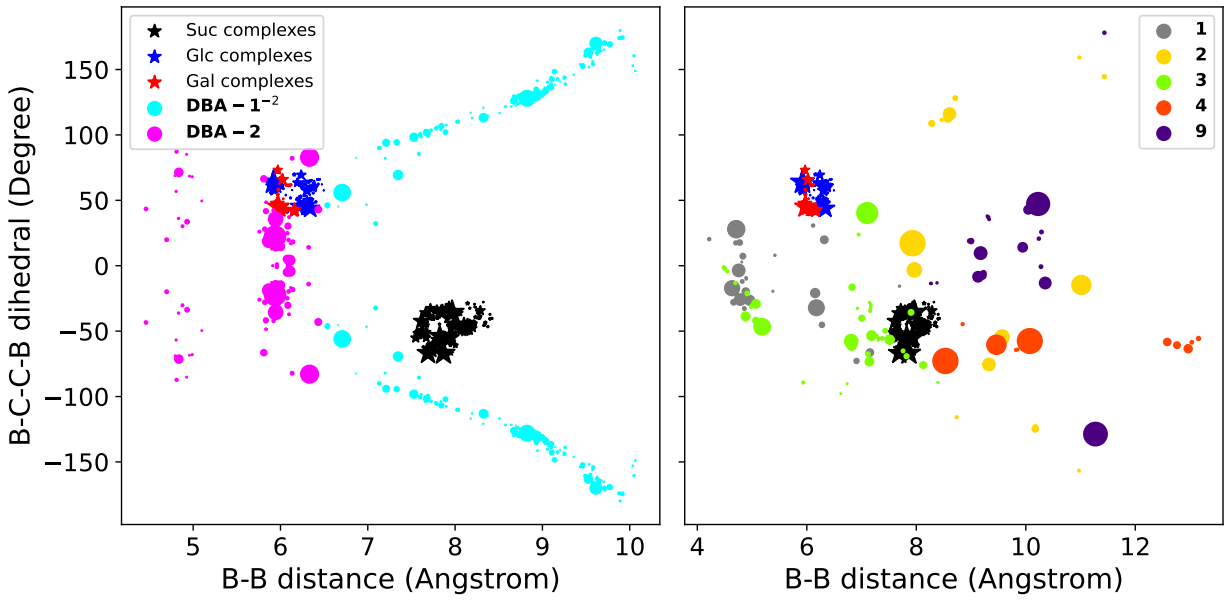


Figure 4.12: Scatter plot of B-B distances (angstroms) and dihedral angles B-C-C-B (degrees) for the free DBA tweezers and their complexes with Suc, Glc, and Gal. The left panel shows the B-B distances and B-C-C-B dihedrals for free tweezers reacting with monosaccharides, while the right panel focuses on tweezers reacting with sucrose, specifically those with $\Delta_r G^\circ_{\mathrm{CE}} < -1.0\,\mathrm{kcal/mol}$, as detailed in Table 4.11. The B-B distances and B-C-C-B dihedrals were calculated from CEs after CENSO refinement. The marker size is proportional to the conformer weight in the Boltzmann distribution. Since **DBA-1**$^{-1}$ and **DBA-2** are symmetric molecules, the negative and positive values of the B-C-C-B dihedrals are equivalent, as illustrated in the left panel.

optimal binding with sucrose, indicating that these configurations are not accessible within the Boltzmann distribution of these tweezers at 298.15 K. In contrast, some dominant conformers of **DBA-1**$^{-1}$ and **DBA-2** are close to the optimal configurations for binding glucose or galactose. These findings are consistent with the observed lack of reactivity between **DBA-1**$^{-1}$ and sucrose.[53,120]

In the case of DBA tweezers for sucrose, a few conformers of tweezer **1** and one dominant conformer of tweezer **3** are close to the optimal configurations for reacting with monosaccharides. It appears that tweezers **1** and **3** are flexible enough to adopt the configurations required for optimal binding to glucose and galactose at 298.15 K. As a result, some interference from these monosaccharides may occur during the quantification of sucrose using DBA tweezers **1** or **3**.

Qualitatively, tweezers **2** and **4** exhibit dominant conformers that are close to the optimal

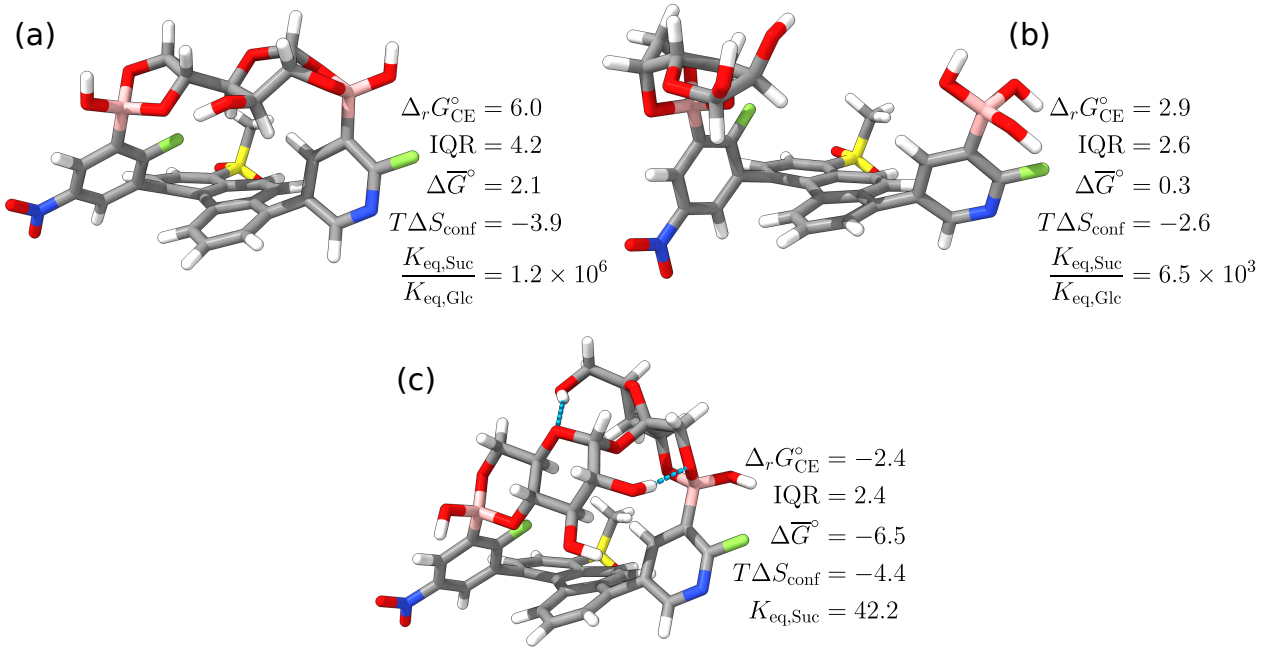


Figure 4.13: Binding motifs of DBA tweezer **4** with glucose (subfigures (a) and (b)) and sucrose (subfigure (c)), alongside the corresponding physicochemical parameters for the reactions. The structure in subfigure (a) resembles the reaction product depicted in Figure 3.1, while the structure in subfigure (b) resembles the reaction product of one aBA reacting with glucose, as shown in Figure 1.6. $\Delta_r G^\circ_{CE}$, IQR, $\Delta\overline{G}^\circ$, $T\Delta S_{conf}$ in kcal/mol. K_{eq} in M^{-1} units.

configuration for sucrose binding, with B-B distances near 8.0 Å and B-C-C-B dihedrals of approximately -60°. In contrast, the dominant conformations of tweezer **9** are observed at B-B distances of 10 Å, indicating that the $R{-}B(OH)_3^-$ groups are predominantly oriented toward the outer part of the tweezer. For tweezer **1**, the most dominant conformers exhibit shorter B-B distances around 5.0 Å. Consequently, in addition to the time required for the reacting molecules to overcome the reaction barrier (as discussed in section 4.4), the reaction of sucrose with tweezers **1** and **9** may be slower compared to tweezers **2** and **4**, since the boronic groups in tweezers **1** and **9** are not pre-aligned for optimal interaction with sucrose. These results suggest that conformational barriers may impede the reaction of DBA tweezers **1** and **9** with sucrose, even though the reactions are exoergic, with $\Delta_r G^\circ_{CE}$ values of -3.2 kcal/mol and -5.1 kcal/mol, respectively.

Our state-of-the-art protocol for calculating reaction free energies[108] ($\Delta_r G^\circ_{CE}$) enables the screening of potential interferences from other carbohydrates due to cross-reactions with the tweezers for sucrose. For example, we estimated the interference of glucose by calculating $\Delta_r G^\circ_{CE}$ for tweezer **4** with two binding motifs of the tweezer **4**-glucose complex, resembling the

structures reported in[50,117] and shown in Figures 1.6 and 3.1. The results of these calculations, presented in Figure 4.13, are compared with the $\Delta_r G^{\circ}_{\mathrm{CE}}$ for the tweezer **4**-sucrose complex.

Reactions of tweezer **4** with glucose are endoergic, despite lower conformational entropy changes ($T\Delta S_{\mathrm{conf}}$) compared to the tweezer**4** + sucrose reaction. The $\Delta\overline{G}^{\circ}$ parameter accounts for the reaction free-energy changes associated with the formation of new B−O−C bonds and intramolecular interactions. For the tweezer**4** + sucrose reaction, the formation of four B−O−C bonds and the optimal positioning of the R−$B(OH)_3^-$ groups overcome the unfavorable entropy changes ($T\Delta S_{\mathrm{conf}}$). In contrast, for the tweezer**4** + glucose reaction, the positive values of $\Delta\overline{G}^{\circ}$ suggest strained structures for the product complexes depicted in Figures 4.13(a) and (b). This strain is exacerbated when four B−O−C bonds are formed, as seen in the product complex shown in Figure 4.13(a).

The equilibrium constants (K_{eq}) for the reactions of tweezer **4** with sucrose and glucose suggest minimal interference from glucose in the quantification of sucrose using tweezer **4**. This is because the double binding of tweezer **4** to sucrose is favored by a factor of 10^6 compared to its double binding to glucose, and by a factor of $\sim 10^4$ in the case of a single R−$B(OH)_3^-$ group of tweezer **4** reacting with glucose.

Based on the findings in this section and the data summarized in Table 4.12, tweezers **2** and **4** emerge as the most promising candidates for implementation in sensors for selective quantification of sucrose. The following sections detail our progress in the chemical synthesis of tweezer **4**, which presents a shorter and more feasible synthetic pathway for its experimental implementation.

4.6 Partial Synthesis of a DBA Tweezer for Selective Detection of Sucrose

The atom backbone of tweezer **4** offers several advantages for the synthesis process: (i) the phenanthrene linker can be synthesized through various reaction mechanisms, providing versatility for incorporating functional groups into this fragment during the synthesis;[187–189] (ii) both aBA moieties are attached to the phenanthrene linker at the 1,3-positions (relative to the boronic group). This arrangement enables the aBAs to be bonded to the phenanthrene linker in a single reaction step,[190] thereby simplifying the synthesis process by reducing complexity and minimizing the number of reaction steps required to purify intermediate products.

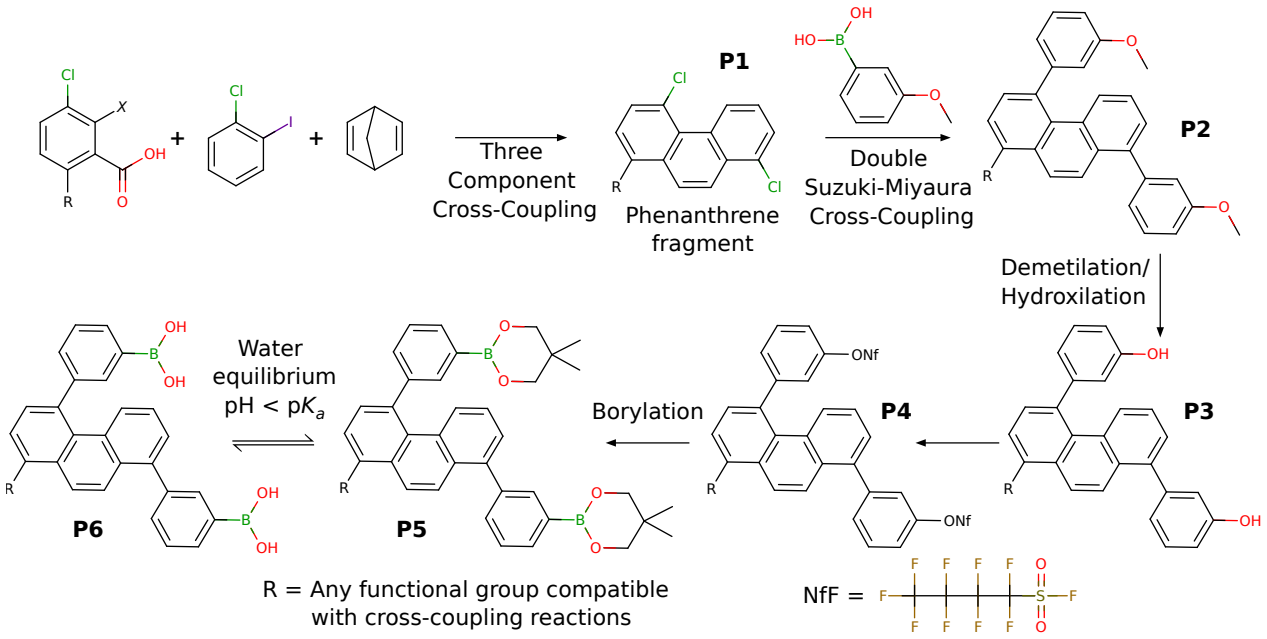


Figure 4.14: Synthetic pathway for compound **P6**, highlighting the aromatic position 'R' in the starting material where functional groups should be introduced to produce a phenanthrene linker with functional groups, similar to the structure in tweezer **4**.

Before proceeding with the synthesis of tweezer **4**, it is common first synthesize a simpler compound to evaluate the synthetic pathway. Tweezer **4** is a derivative of tweezer **22** from Generation 0, as shown in Figure B.1. Therefore, tweezer (0, **22**) was chosen as the core structure to investigate potential synthesis routes and optimize reaction conditions for the subsequent synthesis of tweezer **4**. The synthetic pathway for tweezer (0, **22**). named 1,5-bis(3-boronophenyl)-8-aminophenanthrene, hereafter referred to as **P6**, is outlined in Figure 4.14. The following sections detail the synthesis results for compounds **P1** and **P2**.

4.6.1 Synthesis of 1,5-Diclorophenanthrene (P1)

The 1,5-dichlorophenanthrene (**P1**) serves as the linker fragment to fix the PBA moieties at the precise distance and orientation for reaction with sucrose.To synthesize **P1**, a three-component cross-coupling reaction is proposed,[191–193] as illustrated in Figure 4.15. This reaction utilizes 2-chloro-1-iodobenzene, 2-halo-3-chlorobenzoic acid, and norbornadiene as starting reactants.

The reactant 2-chloro-1-iodobenzene provides the chlorine atom at the 1-position of the phenanthrene linker, while 3-chloro-2-haloarylcarboxylic acid provides the chlorine atom at

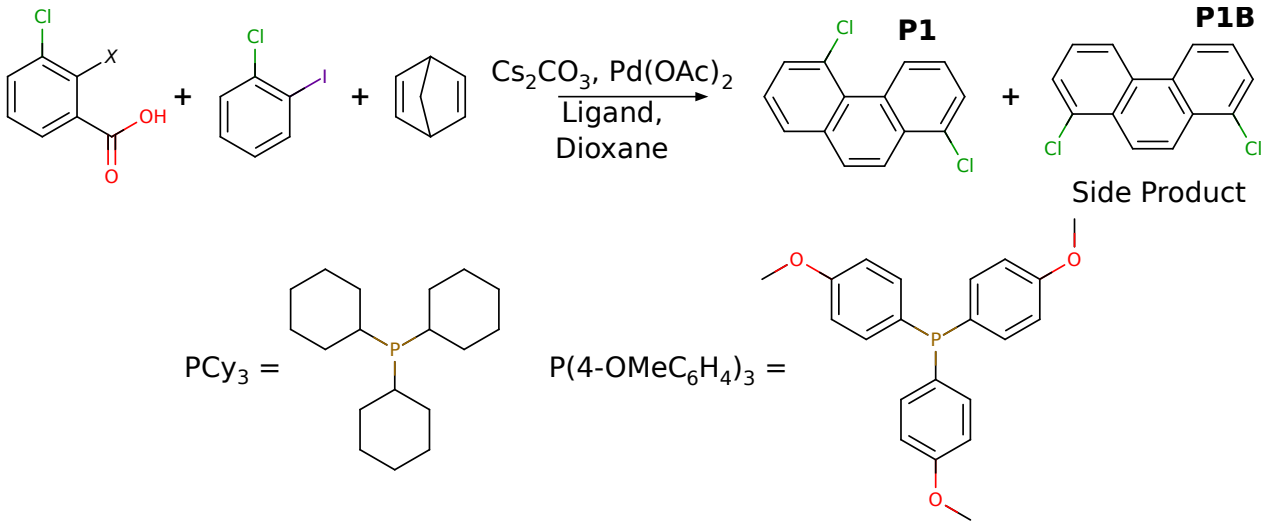


Figure 4.15: One-pot, three-component cross-coupling reaction for the synthesis of compound **P1**. The ligands PCy_3 and $P(4{-}OMeC_6H_4)_3$ were employed to screen the reaction conditions, as summarized in Table 4.12.

the 5-position. The reaction proceeds in a domino fashion within a single reaction flask (one-pot) and involves three sequential cross-coupling steps.[191–193] First, a Catellani reaction couples 1-iodo-2-chlorobenzene and norbornadiene. This is followed by a Suzuki-Miyaura cross-coupling at the 2-position of 2-halo-3-chlorobenzoic acid, linking the three reactants. The reaction concludes with decarboxylation and a retro-Diels-Alder reaction,[191–193] generating **P1**, cyclopentene as a byproduct, and regenerating the Pd catalyst.

From Gas Chromatography-Mass Spectra (GC-MS) experiments, two reaction products with molar masses of 247.12 g/mol were identified, corresponding to the regioisomers **P1** and **P1B**. Analysis of the ^{1}H NMR spectra, shown in Figures D.1 and D.2, confirmed the molecular structures of **P1** and **P1B**, as illustrated in Figure 4.15.

A key distinguishing feature of **P1** is the doublet ^{1}H NMR signal at approximately 9.8 ppm, integrating to one proton, corresponding to the hydrogen at the 6-position (see Figure D.1). This shift arises from the deshielding effect of the chlorine atom at the 5-position in **P1**, which shifts the ^{1}H resonance of the 6-position proton to a higher frequency. In contrast, the ^{1}H signals for the hydrogens at the 5- and 6-positions in **P1B** appear as a combined signal at a lower frequency, around 8.6 ppm, integrating to two protons (see Figure D.2). Additionally, due to the symmetric nature of **P1B**, several ^{1}H signals in the aryl-H resonance region (from 7.5 ppm to 9.0 ppm) overlap. In contrast, the aryl-H resonance region for **P1** appears more crowded due to its asymmetry. These differences in chemical shifts

Table 4.12: Screening of reaction conditions for the one-pot, three-component cross-coupling reaction for the synthesis of compound **P1**. Scale mass unit in mg for 2-halo-1-iodobenzene. Solvent volume in mL.

Entry	X	Ligand	Time	T (°C)	Scale	Solvent	%Yield	**P1**/**P1B**
1	Cl	PCy_3	72 h	130	250	20	–	1:14
2	Cl	PCy_3	18 h	130	250	20	90	1:14
3	Cl	PCy_3	2 h	130	250	20	85	1:14
4	Cl	$P(4-OMeC_6H_4)_3$	18 h	130	250	20	90	1:14
5(x2)	Cl	$P(4-OMeC_6H_4)_3$	18 h	115	50	10	83	1.7:1
6	Cl	$P(4-OMeC_6H_4)_3$	18 h	100	50	10	40	1.7:1
7	Cl	$P(4-OMeC_6H_4)_3$	18 h	115	250	20	92	1:14
8	Br	$P(4-OMeC_6H_4)_3$	18 h	130	250	20	95	1:14

provide clear evidence distinguishing the structures of **P1** and **P1B**.

Table 4.12 summarizes the screening results for the reaction conditions depicted in Figure 4.15. Although high reaction yields were achieved, scaling up the reaction resulted in a significant decline in regioselectivity, with the product ratio shifting from 1.7:1 (**P1**/**P1B**) at 50 mg of 2-halo-1-iodobenzene to 1:14 (**P1**/**P1B**) at 500 mg. This indicates that **P1B** is kinetically favored over **P1**, likely due to steric hindrance in the intermediates leading to **P1**. Similar regioselectivity challenges have been reported for phenanthrene derivatives synthesized from **ortho**-substituted halo-iodobenzenes.[192] The observed regioselectivity ratios for both 2-bromo-1-iodobenzene and 2-chloro-1-iodobenzene at a high mass reaction scale, discard the poor regioselectivity due to the lower reactivity of the aryl-Cl bond compared to the aryl-Br bond in Pd-catalyzed cross-coupling reactions.[194,195]

Since **P1** and **P1B** are regioisomers, their complete separation by column chromatography proves challenging.

4.6.2 Synhtesis of 1,5-Bis(3-methoxyphenyl)phenanthrene (P2)

The next step in the synthetic plan involves attaching 3-methoxyphenyl moieties to compound **P1**, yielding **P2**. These 3-methoxyphenyl groups will act as scaffolds for the subsequent modification into PBA moieties. To streamline the process and minimize intermediate steps, a double cross-coupling reaction is proposed, following a similar one-pot strategy used in the previous stage. This approach aims to eliminate the labor-intensive separation of products associated with sequential single cross-coupling reactions. However, the chlorine atom at

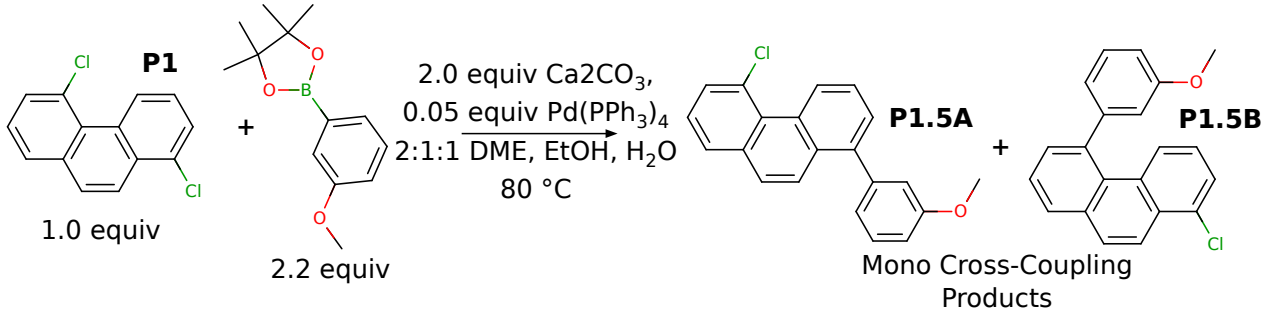


Figure 4.16: Reaction scheme and observed products for the first attempt at synthesizing compound **P2** via a double Suzuki–Miyaura cross-coupling reaction.

the 5-position of **P1** presents significant steric hindrance, complicating the cross-coupling reaction. Consequently, the key challenge at this stage is to identify optimal conditions for the double cross-coupling reaction in a one-pot format.[190,196]

A Suzuki–Miyaura reaction was proposed for reaction of **P1** into **P2**, as illustrated in Figure 4.16, using 3-methoxyphenylboronic acid pinacol ester as the nucleophile.[197] However, the reaction conditions outlined in Figure 4.16 did not proceed to completion. The double substitution of the chlorine atoms with 3-methoxyphenyl groups was not achieved, resulting in a mixture of mono cross-coupling products (**P1.5A** and **P1.5B**) and unreacted starting material. After column purification, 32.3% of the initial reactant **P1** was recovered.

GC-MS analysis revealed a compound with a molecular mass of 318.80 g/mol, consistent with the expected masses of **P1.5A** and **P1.5B**. Furthermore, the ^{1}H NMR spectra (Figure D.3) show a resonance signal at 4.0 ppm, characteristic of protons in alkyl groups, along with aromatic proton (aryl-H) signals. These findings confirm the incorporation of at least one 3-methoxyphenyl group into **P1**.

The incomplete reaction is likely due to reduced reactivity of the remaining chlorine atom in **P1** after the initial coupling step, necessitating a stronger nucleophile to facilitate the second cross-coupling. Additionally, the presence of unreacted starting material suggests a low catalytic turnover rate under the conditions employed (as described in Figure 4.16), which could further hinder the double cross-coupling process.[190]

To address the limitations observed in the reaction conditions shown in Figure 4.16, we replaced the nucleophile with 3-methoxyphenyl boronic acid. PBA are known to be more reactive in the metallation step of the Suzuki–Miyaura coupling.[197,198] A new catalytic system was employed, consisting of $Pd(OAc)_2$ as the palladium source, K_3PO_4 as the base, and SPhos as the supporting ligand.[190,197] The SPhos ligand has demonstrated exceptional efficiency in

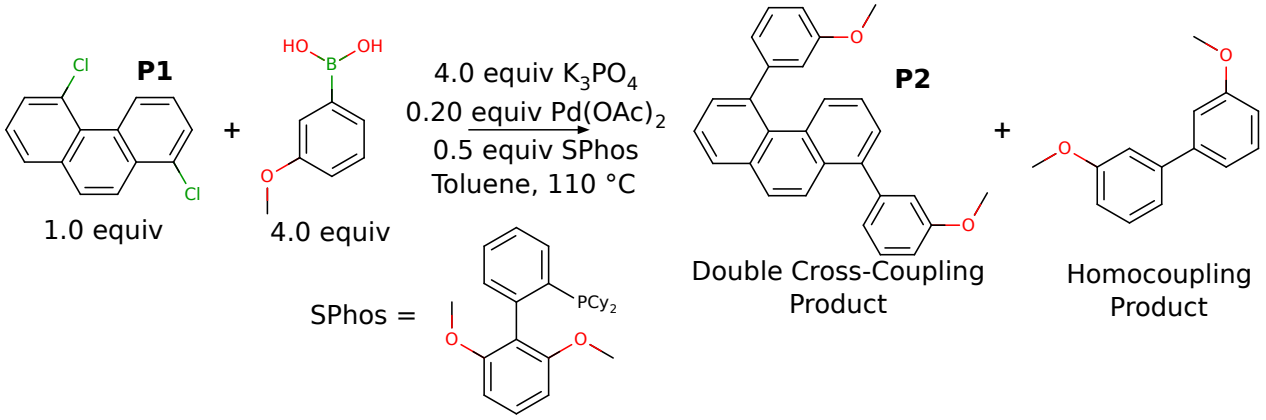


Figure 4.17: Reaction scheme and observed products for the second attempt at synthesizing compound **P2** via a double Suzuki–Miyaura cross-coupling reaction.

the cross-coupling of sterically hindered aryl chlorides and biaryls.[197,199] Additionally, an excess of PBA was used to shift the equilibrium toward product formation.

As a result, the new catalytic system in Figure 4.17 demonstrated excellent activity and high reaction yield for the double cross-coupling of 3-methoxyphenyl boronic acid with **P1**. A product with a molar mass of 390.482 g/mol was identified via GC-MS, consistent with the expected molar mass of **P2**. The structure of **P2** was confirmed through NMR spectroscopy, as shown in Figure D.4. The ^{1}H NMR spectrum revealed key structural changes. The resonance frequency of the proton at the 6-position of the phenanthrene ring in **P2** shifted to a lower frequency, attributable to the removal of the chlorine atom at the 5-position in **P1**. The ^{1}H NMR alkyl signals around 4.0 ppm integrated to six protons, corresponding to the methyl groups. Additionally, the aromatic proton (aryl-H) region became more crowded compared to **P1**. Further evidence of improved catalytic turnover under the new reaction conditions was the observed formation of the homocoupling product of 3-methoxyphenyl boronic acid, the 3,3'-dimethoxy-1,1'-biphenyl, which appeared as a byproduct following the double cross-coupling reaction.

Despite the success of the double cross-coupling in producing **P2**, further optimization of the reaction conditions is required, particularly regarding the amount of 3-methoxyphenyl boronic acid. Using high concentrations of this reactant results in significant formation of the homocoupling byproduct, 3,3'-dimethoxy-1,1'-biphenyl. Traces of this byproduct co-elute with **P2** during chromatographic purification, making the separation process more challenging.

At this stage, we have successfully synthesized the products **P1** and **P2** as outlined in

the synthesis plan in Figure 4.14. However, improving the efficiency of the reaction for **P1** synthesis remains a priority. Compound **P2** has already incorporated the carbon backbone necessary for the final product, **P6**.

Following the reaction plan shown in Figure 4.14, functional groups at the 8-position of the phenanthrene ring (as in tweezer **4**, Figure 4.11) can be selected in the starting materials for first stage of the synthesis plan, provided these groups are compatible with cross-coupling reactions.[190,197] Meanwhile, precursor fragments for aBA moieties with pK_a values below 7.4 can be incorporated during the second reaction stage of the synthesis plan.

The subsequent stages of the synthesis involve converting the methoxy groups ($-OCH_3$) into boronic groups ($-B(OH)_2$) through a series of straightforward reactions. These reactions, well-documented in the literature, typically achieve high yields.[200–202]

4.7 Functionalization of Surfaces for Selective Quantification of Sucrose

In this section, we proposed a protocol for functionalization of a Au(111) surface for sucrose quatification with the DBA tweezer **4** reported in Figure 4.11, Section 4.5.

Our group has expertise in the functionalization of gold nanoparticles for the direct quantification of sugars in real samples, such as fruit juices,[203] as well as in the functionalization of the Working Electrodes (WEs) of Screen-Printed Graphene Electrodes (SPGEs) for the ultra-sensitive quantification of analytes in complex physiological matrices, specifically the detection of the SARS-CoV-2 spike protein.[204] Figure 4.18(b) shows a typical three-terminal SPGE used in our research group for the electrochemical measurement of analytes. The protocols for the functionalization of gold nanoparticles[203] and the WE of SPGE[204] can be adapted, with minimal modifications, for the selective quantification of sucrose using SAMs on Au surfaces. Hereafter, unless otherwise indicated, SPGE refers to a Screen-Printed Gold Electrode.

To ensure compatibility with these protocols and to promote the exposure of the DBA tweezer **4** into the bulk solution above the SAM, the tweezer must be modified replacing the SO_3-CH_3 group to include a primary amino group attached via an alkyl chain of four carbon atoms at the 1-position of the phenanthrene moietie, producing $\mathbf{4}-(CH_2)_4-NH_2$. This configuration is illustrated in Figure 4.18(b).

The synthesis of the modified tweezer $\mathbf{4}-(CH_2)_4-NH_2$ can be achieved following the

strategy outlined in Figure 4.14. The synthetic route begins with the selection of 2-halo-3-chloro-6-R benzoic acid, where R=$-(CH_2)_4-NO_2$, as the starting material (stage 1 in Figure 4.14). The alkyl nitro group (R=$-(CH_2)_4-NO_2$) is well-suited for cross-coupling reactions involving decarboxylation steps,[206] leading to the formation of the intermediate **4**$-(CH_2)_4-NO_2$ in the final stage of the synthetic plan. Subsequently, the aromatic nitro group (ar$-(CH_2)_3-NO_2$) is reduced to an amine group (ar$-(CH_2)_3-NH_2$) using established methods,[207] yielding the final product, **4**$-(CH_2)_4-NH_2$.

The functionalization of the Au surface can be carried out following the protocol described in reference 203. Stock solutions of 0.1 M N–Hydroxysuccinimide (NHS) and 0.1 M N-(3 − Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC·HCl) were prepared separately by dissolving each reagent in a 0.5 M 4-Morpholineethanesulfonic acid (MES) buffer at pH 6.0. These solutions were then mixed in a 3:1 volume ratio, and 100 µL of the resulting NHS:EDC·HCl mixture was added to 10 mL of a freshly prepared 1 mM 4-MBA solution in purified water. This step activated the carboxylic groups of 4-MBA, forming 4-MBA-NHS, which enabled subsequent reaction with **4**$-(CH_2)_4-NH_2$. The SAM coating the Au surface is formed by immersing an etched Au substrate in a 1 mM 4-MBA-NHS solution. The solution is heated to 80 °C and stirred vigorously for 15 minutes to promote the self-assembly of 4-MBA on the Au surface. Usually, this process is repeated twice to ensure complete functionalization. The resulting 4-MBA-NHS SAM is anticipated to form a $(\sqrt{3} \times 4)$ surface lattice,[143,144] as illustrated by the arrangement of sulfur atoms shown in the inset of Figure 4.18(a).

MD simulations, using a canonical NPT ensemble, were performed to calculate the energy of the SAM as a function of the area per tweezer **4**, following the methodology described in Section 3.7 for simulation of the interface water-SAM-Au(111) surface. The DBA tweezers were positioned on the SAM at varying distances, following the $(\sqrt{3} \times 4)$ lattice of 4-MBA on the Au surface. Figure 4.18(a) displays a dissociative curve, indicating a repulsive interaction between the DBA tweezers on the SAM. This repulsion arises from the negative partial charges on the peripheral boronic groups of tweezer **4**, as depicted in Figure 4.18(c). The NPT simulations further revealed tweezer aggregation at dense surface coverages below 720 $Å^2$/tweezer, which is unsuitable for sucrose quantification.

Based on the results of our NPT simulations in Figure 4.18(a), we estimate that preventing tweezer aggregation on the SAM can be achieved by depositing 50 µL of a 0.14 µM solution of **4**$-(CH_2)_4-NH_2$ in Phosphate Buffer Solution (PBS) onto a Au surface with a diameter of 4.4 mm. This ensures the stoichiometric addition of 4.2×10^{12} tweezers for reac-

tion with the activated 4-MBA-NHS in the SAM, corresponding to an Au surface coverage of 720 $Å^2$/tweezer. A reaction time of two hours at room temperature allows the amine group on the DBA tweezer **4**$-(CH_2)_4-NH_2$ to react with the activated carboxylic groups on 4-MBA-NHS, forming a stable peptide bond. The functionalized Au surface with the SAM is then rinsed with PBS to remove unbound reagents and dried at room temperature. Any remaining 4-MBA-NHS can be deactivated using an excess of NH_2-CH_3 in PBS solution. Figure 4.18(b) illustrates the **4**$-(CH_2)_4-NH_2$ tweezer bound to the 4-MBA molecule via a peptide bond ($-C(=O)-NH-$).

Characterization of functionalization can be performed at each step through electrochemical methods, such as cyclic voltammetry and electrochemical impedance spectroscopy, to confirm successful surface modification and monitor changes in the Au surface charge transfer resistance. Additionally, nanoscale morphological analysis of the Au surface can be conducted using atomic force microscopy to ensure uniformity in the functionalization process.[204]

The functionalized Au surface can then be evaluated for sucrose detection. The three-terminal SPGE configuration shown in Figure 4.18(b) can be adapted for sucrose quantification using a FET device. The functionalized WE is modified to function as a metal-oxide gate, with the Au-SAM interface in direct contact with the sample solution. This gate acts as the active layer for detecting and quantifying the target analyte.

Kajisa and co-workers[20] reported the quantification of carbohydrates using an Au surface coated with 4-MBA as the gate-active layer, a doped p-Si semiconductor, and an FET configuration operating in the saturation region. In this region, the drain-source voltage (V_{DS}) exceeds the difference between the gate-source voltage (V_{GS}) and the threshold voltage (V_{Th}), such that $V_{DS} > V_{GS} - V_{Th}$. Under this FET configuration, electrons in the p-Si conduction channel are repelled by the electrostatic field generated by negatively charged complexes $4-\text{ MBA}\cdot\text{sugar}^-$ formed in the reaction $4-\text{MBA} + \text{sugar} \rightleftharpoons 4-\text{ MBA}\cdot\text{sugar}^-$. This repulsion leads to a reduction in the drain-source current (I_{DS}), which is measured as the output potential (V_{out}). Consequently, changes in the output potential (ΔV_{out}) are proportional to the concentration of $4-\text{ MBA}\cdot\text{sugar}^-$ complexes at equilibrium on the Au-SAM interface in contact with the sample solution, allowing the quantification of carbohydrates in solution.[20]

Kajisa and co-workers[20] reported LODs in the micromolar (μM) range for various monosaccharides and disaccharides. However, their system lacked a differential response for mixtures of carbohydrates at concentrations near the LOD, likely due to small differences in the equilibrium constants (K_{eq}) of 4-MBA binding with different carbohydrates. Only fructose and mannose showed a differential response whitin the millimollar (mM) concentration range.

Functionalization of the Au-SAM with DBA tweezer **4** is expected to enhance the differential response of Kajisa's FET sensor towards sucrose. As shown in Figure 4.13, the equilibrium constant (K_{eq}) for the reaction between tweezer **4** and sucrose is approximately four orders of magnitude greater than that for glucose. Consequently, at molar concentrations that maintain a constant chemical potential at the Au-SAM-solution interface,[208] and under near 1:1 concentration ratios in mixtures of sucrose with other carbohydrates, an FET sensor functionalized with DBA tweezer **4** should enable the quantification of sucrose in the sample solution.

The surface coverage of 720 $Å^2$/tweezer and the length of the exposed tweezer on the bulk surface (∼10 Å, see Figure 4.18(b)) correspond to the minimal volume required for the quantification of a single sucrose molecule reacting in a 1:1 ratio with a DBA tweezer. Considering the excluded volume of the exposed tweezer (475.8 $Å^3$), we estimate a theoretical LOD of 0.222 M/k, where k is the number of tweezers on the SAM coating the Au surface. For an Au surface with a diameter of 4.4 mm and a surface coverage of 720 $Å^2$/tweezer, the estimated number of tweezers is 4.2×10^{12}, yielding a theoretical LOD of 52.8 fM. However, the experimental LOD depends on multiple factors, including the sensing device technology, charge screening effects induced by the SAM coating, and the sample matrix, among others. Additionally, optimizing the pH conditions is crucial for achieving high sensitivity and minimal LOD.

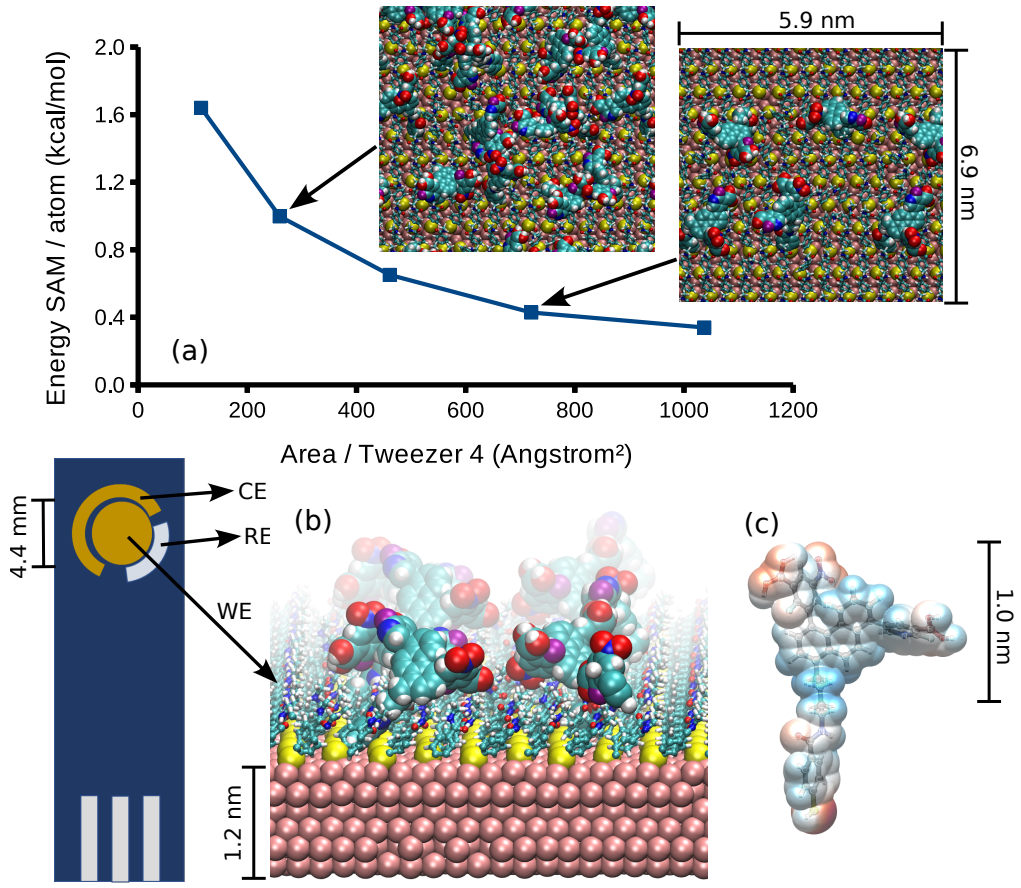


Figure 4.18: Modeling the functionalization of the Au(111) surface with a SAM composed of the tweezer molecule **4** (represented with van der Waals spheres) and 4-mercapto-N-methylbenzamide (ball-and-stick representation) for selective sucrose quantification. (a) Normalized energy of the SAM as a function of the area per tweezer molecule. b) Side view of the functionalized Au(111) surface of a WE in a SPGE, highlighting the exposure of tweezer **4** to the bulk water solution for its reaction with sucrose. RE and CE stand for reference and counter electrodes, respectively. (c) Electrostatic potential isosurface map of the **4** tweezer, with a gradient colors from blue (positive partial charge) through white (neutral) to red (negative partial charge). Partial charges calculated with the electronegativity-equalization method[205] as implemented in Avogadro[145] (version 1.99.0). Sulfur (yellow) and Au (dark pink) atoms in van der Waals representation. Water molecules hidden for better visualization.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

The ability to measure sucrose under real crop-field conditions is critical for breeding plant species with enhanced resilience to biotic and abiotic stresses, especially in climate change caused by global warming. In this work, present several advancements on the design of sucrose sensors the required spatiotemporal resolution for sucrose quantification under physiological solutions or *in vivo* conditions, in real crop-field conditions. Next, we present the achievements related with the specific objective proposed in this work.

Specific Objective 1

Based on our comprehensive literature review, we concluded that tweezer-type diboronic acids exhibit the greatest potential for designing highly selective and biocompatible ligands for the quantification of sucrose in physiological solutions. Their unique structural and functional properties make them ideal candidates for achieving both specificity and compatibility in complex biological environments.

The selection of DBA tweezers was founded on their proven advantages over other molecular recognition methods for carbohydrates: (i) DBA tweezers has demonstrated its biocompatibility and operability for quantification of glucose under *in vivo* conditions. (ii) Compared to enzyme proteins such as INV and SuSy used in biosensors, DBA tweezers are low-molecular-weight molecules, making them more adaptable to different operational conditions. (iii) DBA tweezers exhibit versatility in signal transduction for carbohydrate quantification, supporting their implementation in sensing platforms based on, for example,

fluorescence, spectroscopy, or electrochemical methods. (iv) The modular synthesis approach for DBA tweezers designed for glucose can significantly streamline their adaptation for sucrose.

However, despite the benefits of DBA tweezers for selective glucose quantification under *in vivo* conditions, a clear methodology for extending these advantages to the design of DBA tweezers for sucrose is lacking. This gap in knowledge leads us to Specific Objective 2.

Specific Objective 2

We conducted a comprehensive analysis of the reactivity of sucrose, identifying the most stable products formed through its reaction with aBA molecules. Furthermore, we investigated the reaction mechanism for PBA + sucrose interactions, establishing a foundational framework for the inverse design of DBA tweezer molecules tailored for the selective quantification of sucrose. These insights provide a robust basis for developing highly specific molecular sensors capable of accurately detecting sucrose in complex environments.

Through quantum chemistry calculations, we investigated the stability of the products formed when two aBA molecules —specifically PBA and BOB—in their boronate form $\mathrm{R{-}B(OH)_3}^-$, react with sucrose ($2\,\mathrm{aBA}^- + \mathrm{Suc}$) to produce a cyclic boronate diester compounds, $\mathrm{aBA}^{-1}{-}\mathrm{Suc}{-}\mathrm{aBA}^{-1}$. Stable boronate ester products, characterized by negative reaction enthalpies ($\Delta_r H^\circ$), are form when aBA molecules react with hydroxyl groups at 1',3'-positions, as well as when PBA binds to the 4,6-positions and 2,1'-positions. The cyclic boronate diester compounds formed with $-$OH groups at the 1',3'- and 4,6-positions, which consist of six-membered rings, adopt a stable chair conformation. In contrast, cyclic boronate diesters involving hydroxyl groups at the 1,2-positions were found to be unstable. Although the formation of these boronate diesters introduces chiral centers at the boron atoms, our results do not indicate any preference for a specific chiral configuration in the $\mathrm{aBA}^{-1}{-}\mathrm{Suc}{-}\mathrm{aBA}^{-1}$ products.

The reactivity of sucrose with PBA was analyzed by examining the energy paths for the reaction of PBA with hydroxyls at 1',3'- and 4,6-position of sucrose, under conditions where $\mathrm{pH} < \mathrm{p}K_{a,\mathrm{PBA}}$ and $\mathrm{pH} \geq \mathrm{p}K_{a,\mathrm{PBA}}$, corresponding to sucrose reacting with PBA in its trigonal ($\mathrm{R{-}B(OH)_2}$) and tetrahedral ($\mathrm{R{-}B(OH)_3}^{-1}$) forms, respectively. Our results confirm that boronate PBA^{-1} ($\mathrm{R{-}B(OH)_3}^{-1}$) is the most reactive species toward sucrose, exhibiting reaction barriers below 20 kcal/mol and exoergic reaction-free energies. Under standard conditions, the reaction of PBA to hydroxyls at the 4,6-positions is predicted to occur within minutes, while reaction at the 1',3'-positions is expected to require at least two hours. Rais-

ing the reaction temperature to 40 °C is anticipated to accelerate the kinetics, reducing the timescale for both reactions to within minutes. In contrast, reactions involving trigonal PBA ($R-B(OH)_2$) with sucrose are endoergic, with reaction half-life times at standard conditions too slow for practical continuous sucrose monitoring. Our results suggest that sucrose quantification with DBA tweezer molecules should be performed at reaction conditions $pH \geq pK_{a,PBA}$. However, further studies are needed to explore the potential catalytic effects of water molecules on the reactivity of $R-B(OH)_2$. Moreover, additional information—such as the limits of detection and quantification—is required to evaluate whether reaction half-lives on the scale of minutes or hours are sufficient for continuous sucrose monitoring in plants. These considerations are highly dependent on the efficiency of the transduction signal used to detect the DBA tweezer-sucrose product.

The stable products $aBA^{-1}-Suc-aBA^{-1}$ were used as scaffolds to design DBA tweezers with the optimal distance and orientation for reaction with sucrose. Among the configurations, $PBA(S)^{-1}-Suc-PBA(S)^{-1}$ (where *"S"* denotes the chirality of the boron atoms) generated the largest number of unique DBA tweezers (1361). In this configuration, the phenyl rings of both PBA units adopt an almost parallel (*endo-*) arrangement, providing the shortest distance between the two PBA moieties and facilitating the incorporation of linker molecular fragments. In contrast, other configurations produced fewer than 10 unique DBA tweezers, as only a limited number of molecular linkers in the HostDesigner library possessed the appropriate length to close the tweezer structure. Furthermore, none of the $BOB^{-1}-Suc-BOB^{-1}$ scaffolds yielded DBA tweezers, likely due to the inherent rigidity of the BOB moiety. The use of generative models or the inclusion of alternative linker fragment libraries could potentially expand the diversity of unique DBA tweezers based on scaffolds other than $PBA(S)^{-1}-Suc-PBA(S)^{-1}$. However, these approaches might result in tweezers with higher molecular weights, potentially increasing the complexity of their synthesis.

At this stage, we have successfully designed a group of 1,361 DBA tweezers with optimally positioned boronic groups for reaction with sucrose. However, these tweezers are anticipated to have low water solubility due to their predominantly carbon-based composition. Furthermore, it remains uncertain whether the reaction between the DBA tweezers and sucrose proceeds spontaneously in aqueous media. These critical challenges were the focus of Specific Objective 3.

Specific Objective 3

In this work, we developed a computational workflow for the inverse design of DBA tweezers tailored for the highly selective quantification of sucrose. This approach enables the design of tweezers with optimized physicochemical parameters, ensuring effective sucrose quantification in aqueous solutions. Our methodology leverages a state-of-the-art, semi-automated workflow grounded in a robust physical framework for calculating the absolute reaction-free energies of DBA tweezers reacting with carbohydrates in aqueous environments. This advancement provides a reliable and efficient pathway for designing selective molecular sensors with tailored binding properties.

The accuracy of our workflow was validated against experimental results for DBA tweezers reacting with glucose and galactose. The strong interactions between DBA tweezers and water molecules necessitated the microsolvation to model solute-solvent interactions accurately. Additionally, the (unfavorable) conformational entropy changes associated with monosaccharide-tweezer complexation contributed approximately 3–5 kcal/mol to the reaction-free energy. After fine-tuning the hyperparameters, our workflow achieved a mean absolute error of 2.2 kcal/mol with a variance, measured as the interquartile range, of 2.4 kcal/mol. This level of accuracy enables the identification of tweezer-carbohydrate reactions that are thermodynamically favorable under standard conditions. However, the high computational cost of the workflow, in its current form, limits its application for large-scale high-throughput screening.

The tweezer design process for sucrose was completed with the implementation of a GB-GA method for holistic, multi-objective optimization of DBA tweezers designed to react with sucrose in aqueous media at physiological pH (7.4). The multi-objective optimization incorporated key parameters into the fitness function, including reaction free energy, $\log P_{\mathrm{o/w}}$ for water solubility, the maximum pK_a of the aryl boronic moiety, and SA score. A representative subset of the 1361 tweezers designed with the PBA$(S)^{-1}-$Suc$-$PBA$(S)^{-1}$ scaffold was selected as Generation 0. Fine-tuning parametrization of the GB-GA balanced the exploration and exploitation of the chemical design space. The optimization process led to substantial improvements across several properties of the DBA tweezers. Enhanced designs exhibited improved water solubility, pK_a values compatible with the reaction of aBAs with sucrose at physiological conditions, and novel structural features absent from the initial population. However, the SA score increased during the optimization process, suggesting potential challenges in practical synthesis. Addressing this issue in future work may involve employing alternative scalarization functions or adjusting weighting parameters within the

multi-objective optimization framework to prioritize simpler designs.

High-accuracy calculations of reaction-free energies, conducted after multi-objective optimization, confirmed that the inclusion of conformational entropy favored rigid tweezer designs. These rigid tweezer structures effectively surmount the unfavorable conformational entropy changes for tweezer-sucrose complexation, by favorable covalent and non-covalent interactions, enabling spontaneous reactions despite sucrose's inherent flexibility. A boron-boron distance of approximately 7.9 Å and dihedral angles between boronic groups near -54° were identified as critical for high selectivity toward sucrose. Notably, the reaction-free energies for some DBAtweezer + sucrose complexations were comparable to those calculated for monosaccharides, demonstrating the effectiveness of our approach in designing DBA tweezers for the more flexible carbohydrate sucrose.

The selectivity analysis was conducted by calculating the free energies and equilibrium constants for the cross-reaction of the DBA tweezer **4** with glucose. The reactions between tweezer **4** and glucose were found to be endoergonic. Notably, the equilibrium constant for the tweezer**4** + sucrose reaction was six orders of magnitude greater than that for the double bonding of tweezer **4** with glucose, and approximately four orders of magnitude higher than the reaction involving a single boronic group of tweezer **4** with glucose. These results demonstrate that our computational design successfully achieves a highly selective DBA tweezer for sucrose.

Guided by selectivity analysis and chemical insights, the top-performing DBA tweezer **4**, presented in Table 4.11, was chosen for synthesis as outlined in Specific Objective 4.

Specific Objective 4

We proposed a synthesis plan for sucrose-specific DBA tweezers featuring a phenanthrene linker substituted with two 3-phenyl boronic acid groups at the 1,5-positions of the phenanthrene linker.

The first stage of the synthesis employed a domino reaction consisting of a three-component cross-coupling to produce 1,5-dichlorophenanthrene. The target product was successfully confirmed using GC-MS and ^{1}H NMR experiments, and the reaction achieved a high yield. However, the process exhibited poor regioselectivity at reaction scales exceeding 50 mg of the limiting reactant, resulting in 1,8-dichlorophenanthrene as the major product. In the second stage, 1,5-bis(3-methoxyphenyl)phenanthrene was synthesized via a Suzuki-Miyaura double cross-coupling reaction between 1,5-dichlorophenanthrene and 3-methoxyaryl boronic acids. After optimizing the reaction conditions, the desired product was successfully ob-

tained, as confirmed by GC-MS and ^{1}H NMR experiments. This achievement represents the successful assembly of the atomic backbone required for the DBA tweezer, 1,5-bis(3-boronophenyl)phenanthrene. Importantly, the versatility of our synthesis plan allows for the incorporation of functional groups into the phenanthrene linker during the first synthesis stage to enhance water solubility. Additionally, aBAs with pK_a values below 7.4 can be integrated into the tweezer during the second synthesis stage. This modular approach underscores the synthetic accessibility of DBA tweezers specifically designed for selective sucrose detection in aqueous environments, as conceived using the GB-GA multi-objective optimization developed in this work.

The low regioselectivity observed during the synthesis of 1,5-dichlorophenanthrene posed significant challenges in completing the proposed synthesis plan for 1,5-bis(3-boronophenyl) phenanthrene. This compound was intended as a pilot molecule to optimize the synthetic strategy for the subsequent preparation of the DBA tweezer **4**, designed through the GB-GA optimization process. Efforts to identify optimal conditions for improving the regioselectivity of 1,5-dichlorophenanthrene synthesis consumed the majority of the four months allocated to finalizing the synthetic route for 1,5-bis(3-boronophenyl)phenanthrene.

The successful synthesis of the carbon backbone for 1,5-bis(3-boronophenyl)phenanthrene enabled us to propose a plausible DBA tweezer for functionalizing a SPGE for selective sucrose quantification, contributing to the fulfillment of Specific Objective 5.

Specific Objective 5

We proposed the design of a FET device featuring a SAM coating on an Au surface, utilizing the DBA tweezer **4** and 4-MBA molecules. This device is anticipated to enable the highly selective quantification of sucrose, with a signal intensity for sucrose detection six orders of magnitude greater than that for glucose. Additionally, the device demonstrates an upper limit of detection (LOD) in the femtomolar range.

The synthesis plan for the DBA tweezer 4 enables its modification to incorporate an alkyl amino chain at the phenanthrene moiety, yielding **4**$-(CH_2)_4-NH_2$, a compound suitable for SAM coating on Au surfaces. This modification ensures that the boronic acid groups are exposed to the bulk solution, allowing efficient interaction with sucrose. The SAM functionalization of Au surfaces can be performed using the established protocol developed by our research group, with its progress monitored through electrochemical characterization and atomic force microscopy to verify proper electrode surface modification.

NPT simulations indicate that an Au surface coverage of 720 $Å^2$/tweezer prevents aggrega-

tion of **4**$-(CH_2)_4-NH_2$ within the SAM layer, providing an optimal configuration for sucrose quantification. On an Au surface with a diameter of 4.4 mm, this surface coverage enables sucrose detection at femtomolar concentrations under optimal FET device and measurement conditions. In the presence of molar concentrations maintaining a constant chemical potential at the Au-solution interface[208] and near 1:1 sucrose-to-carbohydrate ratios, an FET sensor functionalized with DBA tweezer 4 should enable sucrose quantification.

The quantitative correlation between surface charge density changes and sucrose concentration enables precise and reliable detection using FET-based sensing technologies. Moreover, this approach can be easily adapted for detecting other carbohydrates by simply modifying the tweezer molecules, establishing a versatile platform for ultrasensitive and selective carbohydrate detection in complex environments with direct applicability to advanced biosensing.

Additonal Outcomes

In addition to achieving the specific objectives outlined for this work, two significant additional outcomes emerged during the investigation:

1. State-of-the-Art Workflow for Reaction-Free Energy Calculations: A workflow grounded in robust physicochemical principles for accurate calculation reaction-free energies between DBA tweezers and carbohydrates was developed and published in reference 108.

2. Automated Multi-Objective Optimization Workflow: An automated workflow for the multi-objective optimization of DBA tweezers was implemented. While initially applied to design DBA tweezers for sucrose, this program is adaptable for designing DBA tweezers for any monosaccharide or disaccharide. This is achieved by modifying the core fragment $aBA^{-1}-sugar-aBA^{-1}$ and adjusting the initial population of DBA tweezers accordingly.

5.2 Future Work

The findings of this work open up several promising lines for future research. Below, we outline some of the research projects that merit further exploration.

Future work 1: The next logical step in this research is the experimental validation of product formation for the reaction between the DBA tweezers and sucrose, with a particular

focus on confirming the cyclic boron diester structure involving the hydroxyl groups at the 1',3'-positions of the fructose moiety. Achieving this goal necessitates the complete synthesis of one of the top-performing tweezers.

The synthesis of a DBA tweezer for sucrose will enable validation of the reaction mechanism for DBA-tweezer formation, optimization of conditions for sucrose quantification in various samples, and exploration of different sensing platforms and transduction mechanisms for detecting the DBA-sucrose complex. These objectives are well-suited for a master's research plan.

Expanding this scope to include the quantification of other carbohydrates, such as trehalose, would help validate the generality of the design protocol proposed in this work. A doctoral research plan would be more appropriate for the synthesis of two DBA tweezers for the selective quantification of different carbohydrates, the implementation of these tweezers on a sensing platform, and the optimization of sensing conditions, as it would demonstrate a broadly applicable method for the specific quantification of carbohydrates in aqueous solutions.

Future work 2: Other branch for this work may focus on improving the performance and thoroughput of the protocol proposed in this work.

A critical factor in improving the performance of the GB-GA algorithm is increasing the throughput of reaction-free energy calculations while maintaining a mean absolute error near the threshold of chemical accuracy. Higher throughput enables extensive hyperparameter optimization of the GB-GA algorithm, thereby enhancing its overall performance and expanding the explored chemical space by allowing larger population sizes. To achieve this, we are parameterizing a reactive force field (ReaxFF) to accurately model the reactivity of boronic and boronate compounds with sucrose in aqueous media.

However, as demonstrated in this work, achieving highly accurate calculations of reaction free energies for DBA tweezers with carbohydrates relies on accurately sorting conformational ensembles, which in turn depends on the precise computation of intramolecular energies—currently achievable only through quantum chemistry methods. Using a force field like ReaxFF to sort conformational ensembles of DBA tweezers, carbohydrates, and their reaction products in solution would require extensive validation and potential reparameterization of the force field. Alternatively, machine-learning approaches have demonstrated the accuracy of *ab initio* quantum chemistry methods for evaluating molecular PES. However, developing machine-learning-based PES models requires a vast amount of training data, which is currently scarce for DBA-carbohydrate systems and must be "synthetically" generated using a

cost-efficient approach.

This approach aims to streamline reaction-free energy calculations, enabling more efficient algorithm training and broader exploration of the chemical space using the GB-GA algorithm. These improvements to the *in silico* design proposed in this work could form the basis of a master's thesis research plan.

Nevertheless, the design process of DBA tweezers can be further improved by incorporating additional optimization strategies. The multi-objective optimization framework provides the flexibility to integrate new parameters into the design process. For instance, features related to signal transduction could be integrated into the optimization process, or a reward function could be implemented to favor the selection of free tweezers with boronic groups pre-coordinated for sucrose reaction. Furthermore, adopting vectorial multi-objective optimization schemes, such as those based on the Pareto method, could enhance the optimization process by achieving a more effective balance among conflicting parameters.

The reaction mechanism studied for PBA and sucrose can be extended to investigate the reactions between boronic acids and diols. By coupling the reaction pathways of $R{-}B(OH)_2$ and $R{-}B(OH)_3^-$ with diols and analyzing the acid-base equilibria of all reactants and products, valuable insights can be gained into the reaction kinetics and thermodynamics of boronic acid–diol systems, which depend on initial pH conditions. Using *ab initio* calculations, this understanding can be applied to optimize the performance of boronic acid-based sensors. Specifically, these calculations can guide adjustments to the sample pH or the selection of aryl boronic acids with pK_a values tailored to the solution's pH. Such optimization would enhance the stability constant of the cyclic boron-diester product, ensuring optimal performance of a sensor under targeted conditions.

All the proposed improvements to the DBA tweezer design protocol outlined in this work could form the basis of a doctoral thesis.

Future work 3: Our protocol for designing DBA tweezers can be readily adapted to target other carbohydrates, including monosaccharides and disaccharides, by modifying the core scaffold for the design process using HostDesigner and the optimization process with the GB-GA algorithm. However, this adaptation requires precise characterization of the reaction products formed between the carbohydrate and aryl boronic acids. Extending the approach to more complex systems, such as polysaccharides and polyphenol molecules, is also feasible. This would likely necessitate incorporating machine-learning generative methods to design linker molecules, along with improvements in throughput for reaction-free energy calculations to ensure computational efficiency.

In future work, we plan to improve the overall performance of the DBA tweezers by the design of new hybrid DBA tweezer-temple receptors. Temple receptors enhance selectivity by screening carbohydrates based on their shape and size, and increase affinity toward carbohydrates through hydrogen bonds, C-H $\cdots \pi$ and polar interactions. They suffer however from low signal transduction. The hybrid systems will offer the advantages of both receptor types, i.e. high selectivity of temple receptors and strong signal of DBA-carbohydrate complexes, expanding the spectrum of available detection techniques.

Due to the high-risk, high-reward nature of this proposed work, its scope may necessitate the involvement of a master's student, contributing to and supporting the development of a broader doctoral thesis.

Appendix A

Supplementary Data for Free Energy Calculations of DBA Reactions with Carbohydrates

A.1 Calculation of Molecular Free Energies $G^{\circ}(X)$ with the Cluster-Continuum Approximation

Under a cluster-continuum approximation to model the aqueous (aq) environment, the molecular free energy G° for a solute X surrounded by n water molecules is given by the eq A.1:[81]

$$G^{\circ}_{aq,n}(X) = E_{\text{gas}}(X) + G^{\circ}_{\text{trv}}(X) + \delta G^{\circ}_{\text{solv},n}(X), \tag{A.1}$$

where $E^{\circ}_{\text{gas}}(X)$ is the electronic energy for the solute X in the gas phase, $G^{\circ}_{\text{trv}}(X)$ is the free energy resulting from the thermostatistical contributions of the molecule's translations, rotations, and vibrations, and $\delta G^{\circ}_{\text{solv},n}(X)$ is the solvation free energy for the solute surrounded by the n water molecules. Bryantsev[80] numerical findings demonstrate that a thermodynamic cycle employing water clusters for calculation of $G^{\circ}_{solv,n}(X)$ exhibits a gradual convergence with increasing the water cluster size (n).[80,81] According to Jensen,[81] using the cluster approach from Bryantsev[80] for calculation of $G^{\circ}_{solv,n}(X)$, the expanded version of eq A.1 becomes:

$$\begin{aligned} G^{\circ}_{aq,n}(X) = {} & E_{\text{gas}}(X(\text{H}_2\text{O})_n) + G^{\circ}_{\text{trv}}(X(\text{H}_2\text{O})_n) + \delta G^{\circ}_{\text{solv}}(X(\text{H}_2\text{O})_n) - RT\ln(n!) \\ & - E_{\text{gas}}((\text{H}_2\text{O})_n) - G^{\circ}_{\text{gas,trv}}((\text{H}_2\text{O})_n) - \delta G^{\circ}_{\text{solv}}((\text{H}_2\text{O})_n) - RT\ln([\text{H}_2\text{O}]/n). \end{aligned} \tag{A.2}$$

$RT\ln(n!)$ is a factor to account for the permutation of water, and $RT\ln([H_2O]/n)$ is a correction term for the concentration change of the n water molecules from 1 M to the standard water concentration ($[H_2O]$ = 55.34 M). Calculating the molecular free energy ($E_{gas}(X(H_2O)_n) + G^{\circ}_{trv}(X(H_2O)_n) + \delta G^{\circ}_{solv}(X(H_2O)_n)$) from the average over a CE, the eq A.2 turn into:

$$G^{\circ}_{CE}(X) = G^{\circ}_{CE,aq}(X(H_2O)_n) - G^{\circ}_{CE,aq}((H_2O)_n) - RT\ln(n!) - RT\ln([H_2O]/n). \quad (A.3)$$

The terms $G^{\circ}_{CE,aq}(X(H_2O)_n)$ and $G^{\circ}_{CE,aq}((H_2O)_n)$ are given directly from the output of CENSO.

We used the eq A.4 to calculate the molecular free energy for the water molecules ($(H_2O)_4$) produced after the reaction of the DBAs with the carbohydrates (Figures 3.1, 3.2 and 3.3):

$$G^{\circ}_{CE}(H_2O)_4 = G^{\circ}_{CE,aq}((H_2O)_{4+n}) - G^{\circ}_{CE,aq}((H_2O)_n) - RT\ln(n!) - RT\ln([H_2O]/(4+n)). \quad (A.4)$$

In this case, the four water molecules from the reaction should not be included in the term $RT\ln(n!)$ for permutation of water molecules but were included in the correction for the normal concentration of water $RT\ln([H_2O]/(n+4))$.

The molecular free energies $G^{\circ}_{CE,aq}(X(H_2O)_n)$ and $G^{\circ}_{CE,aq}((H_2O)_n)$ in eq A.3 should be sampled several (N) times because of the stochastic process for the generation of the CE. Therefore, the selection of any particular combination of trials $G^{\circ}_{CE,\alpha}(X(H_2O)_n)$[1] and $G^{\circ}_{CE,\beta}((H_2O)_n)$[1], with $\alpha, \beta \in \{1, 2, 3, ..., M\}$, is also arbitrary. In analogy to eq 3.2, we defined $G^{\circ}_{CE}(X)$ as the set of all the possible permutation for α, β indices, turning eq A.3 into eq A.5:

$$\begin{aligned} G^{\circ}_{CE}(X) &= \{G^{\circ}_{CE,i}(X)\} \\ &= \{G^{\circ}_{CE,\alpha}(X(H_2O)_n) - G^{\circ}_{CE,\beta}((H_2O)_n) - RT\ln(n!) - RT\ln([H_2O]/n)\}, \end{aligned} \quad (A.5)$$

where a particular combination of $\{\alpha, \beta\}$ corresponds to the index $i \in \{1, 2, 3, ..., M^2\}$. Equivalently, eq A.4 turns into eq A.6:

$$\begin{aligned} G^{\circ}_{CE}(H_2O)_4 &= \{G^{\circ}_{CE,i}((H_2O)_4)\} \\ &= \{G^{\circ}_{CE,\alpha}((H_2O)_{4+n}) - G^{\circ}_{CE,\beta}((H_2O)_n) - RT\ln((n-4)!) - RT\ln([H_2O]/(4+n))\}. \end{aligned} \quad (A.6)$$

[1]The symbol *aq* to denote aqueous media was removed to avoid the over-crowding of subindices

For the calculation of $\Delta_r G^{\circ}_{\mathrm{CE}}$, the terms $G^{\circ}_{\mathrm{CE},aq}((\mathrm{H_2O})_n)$ of the reactants cancel out with the ones of the products when the same number of water molecules are used for the for micro-solvation of the reactants and products. For example, in a reaction of the type $A(\mathrm{H_2O})_n + B(\mathrm{H_2O})_n \rightleftharpoons C(\mathrm{H_2O})_n + D(\mathrm{H_2O})_n$.

A.2 Equilibrium Reactions

A.2.1 Speciation Digrams

The reactions of both **DBA-1**$^{-2}$ and **DBA-2** with glucose used for benchmark in this work are represented by the next equations:

$$\textbf{DBA-1}^{-2} + \mathrm{Glc} \overset{K}{\rightleftharpoons} \textbf{DBA-1.Glc}^{-2} + 4\,\mathrm{H_2O}, \tag{A.7}$$

$$\textbf{DBA-2} + \mathrm{Glc} \overset{K}{\rightleftharpoons} \textbf{DBA-2.Glc} + 4\,\mathrm{H_2O}. \tag{A.8}$$

However, there are other simultaneous equilibrium reactions for the reactants and products of reactions A.7 and A.8 that may be important depending on the reaction pH's. For reaction A.7, the mentioned equilibrium reactions are given by the following equations:

$$\textbf{DBA-1} \overset{K_{a1}}{\rightleftharpoons} \textbf{DBA-1}^{-} + \mathrm{H^+} \overset{K_{a2}}{\rightleftharpoons} \textbf{DBA-1}^{-2} + \mathrm{H^+}, \tag{A.9}$$

$$\textbf{DBA-1.Glc} \overset{K_{a1}}{\rightleftharpoons} \textbf{DBA-1.Glc}^{-} + \mathrm{H^+} \overset{K_{a2}}{\rightleftharpoons} \textbf{DBA-1.Glc}^{-2} + \mathrm{H^+}. \tag{A.10}$$

For reaction A.8, the other equilibrium reactions are given by the following equations:

$$\textbf{DBA-2}^{+2} \overset{K_{a1}}{\rightleftharpoons} \textbf{DBA-2}^{+} + \mathrm{H^+} \overset{K_{a2}}{\rightleftharpoons} \textbf{DBA-2} + \mathrm{H^+} \tag{A.11}$$

$$\textbf{DBA-2.Glc}^{+2} \overset{K_{a1}}{\rightleftharpoons} \textbf{DBA-2.Glc}^{+} + \mathrm{H^+} \overset{K_{a2}}{\rightleftharpoons} \textbf{DBA-2.Glc} + \mathrm{H^+}. \tag{A.12}$$

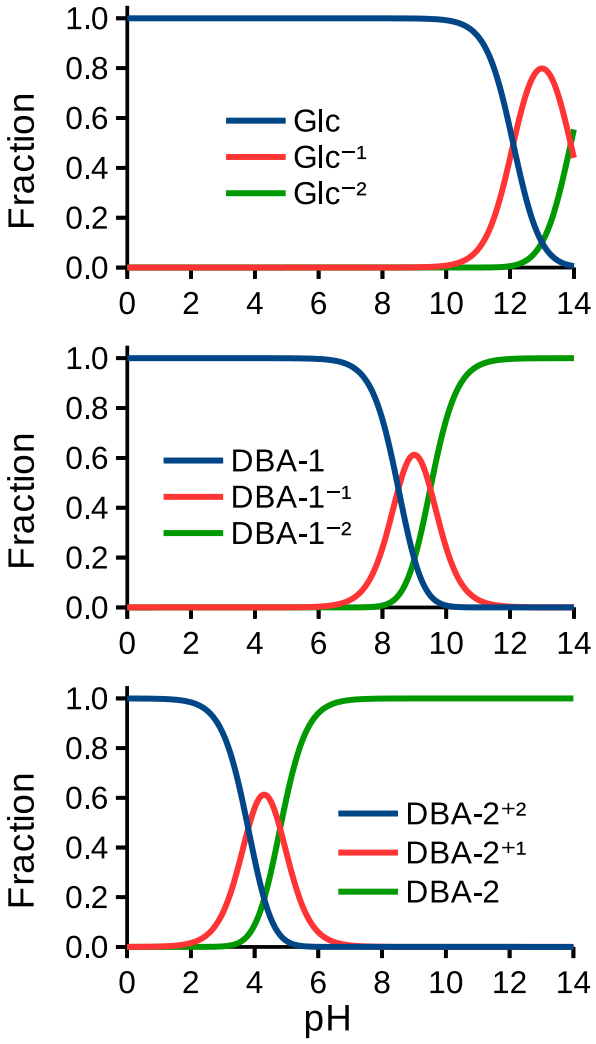


Figure A.1: Speciation diagrams as a function of the pH for Glc, **DBA-1**, and **DBA-2**.

For both reactions A.7 and A.8, the equilibrium for glucose (eq A.13) may also be present:

$$\mathrm{Glc} \overset{K_{a1}}{\rightleftharpoons} \mathrm{Glc}^{-} + \mathrm{H}^{+} \overset{K_{a2}}{\rightleftharpoons} \mathrm{Glc}^{-2} + \mathrm{H}^{+} \tag{A.13}$$

We present the speciation diagrams for glucose, **DBA-1**, and **DBA-2** in Figure A.1. For **DBA-2**, the pK_as reported correspond to pK_{a1} = 3.7 and pK_{a2} = 4.7.[55] The pK_{a2} is close to the acidity constant of the parent compound N-methyl-3-pyridiniumboronic acid (pK_a = 4.4), and the pK_{a1} is almost one unit less than the pK_a of the parent compound. Although there are no reports for the pK_as of **DBA-1**, to get the speciation diagram we assumed a value of pK_{a2} = 9.5, close to the pKa of the parent compound *p*-tolylboronic acid

($pK_a = 9.3$),[117] and a value of $pK_{a1} = 8.5$, one unit less than pK_{a2} (in analogy to **DBA-2**). The pK_as reported for glucose are $pK_{a1} = 12.1$ and $pK_{a1} = 13.9$.[118,119]

A.2.2 Equilibrium reactions for the DBA-1^{-2} molecule

The reaction A.7 was performed at pH = 11.3.[53,117,120] At this pH, we reported in Table 1 a fraction $\alpha = 0.88$ for glucose in equilibrium with deprotonated glucose Glc^{-1} (see Figure A.1(a)). The double-deprotonated glucose (Glc^{-2}) is not present in the solution. To quantify the magnitude of equilibrium $Glc \rightleftharpoons Glc^{-1} + H^+$ (eq A.13) in the calculation of the free energy $\Delta_r G^\circ_{CE}$ for reaction A.7, we used the conditional equilibrium constant K' given by the eq A.14:[46]

$$K' = \frac{[\textbf{DBA-1.Glc}^{-2}]}{[Glc'][\textbf{DBA-1}^{-2}]} = \frac{[\textbf{DBA-1.Glc}^{-2}]}{[Glc][\textbf{DBA-1}^{-2}]}\,\alpha_{Glc} = K\,\alpha_{Glc}, \tag{A.14}$$

where [Glc′] is the total concentration of glucose given by the sum of neutral ([Glc]) and deprotonated glucose [Glc^{-1}], and α_{Glc} is the side-reaction coefficient for glucose. Glucose concentrations and α_{Glc} are related by eq A.15:

$$[Glc'] = [Glc] + [Glc^{-1}] = [Glc]\left(1 + \frac{K_{a1}}{[H^+]}\right) = \frac{[Glc]}{\alpha_{Glc}}. \tag{A.15}$$

Replacing eq A.14 into eq 1 yields eq A.16:

$$\begin{aligned}\Delta_r G^\circ_{Exp} = -RT\ln(K\,\alpha_{Glc}) = -RT\ln(K) - RT\ln(\alpha_{Glc}) &= \Delta_r G^\circ_{CE} - RT\ln(\alpha_{Glc}) \\ &= \Delta_r G^\circ_{CE} + 0.07\,\text{kcal/mol},\end{aligned} \tag{A.16}$$

where $\Delta_r G^\circ_{CE}$ is the reaction-free energy calculated with our CREST-CENSO workflow for the reaction A.7. At pH = 11.3, the value of $\alpha_{Glc} = 0.88$ (see Table 3.1). Therefore, the last term of eq A.16 $-RT\ln(\alpha_{Glc}) = 0.07$ kcal/mol, which allows us to conclude that ignoring the equilibrium A.13 does not introduce a significative error source for the calculation of the equilibrium constant K (and the reaction-free energy $\Delta_r G^\circ$) for the reaction A.7.

A.2.3 Equilibrium Reactions for the DBA-2 Molecule

The reaction A.8 was perform at pH = 7.4.[55] The speciation diagrams in Figures A.1(a) and A.1(c) show that at pH = 7.4 for reactants the only species present are the zwitterionic **DBA-**

2 and the neutral glucose, as reported in Table 3.1. Therefore, the equilibrium reactions A.11 and A.13 are not important at pH = 7.4 and can be ignored for the calculation of the equilibrium constant for reaction A.8.

A.2.4 Equilibrium Reactions for DBA-1.Glc^{-2} and DBA-2.Glc Molecules

We showed in the previous sections that only the **DBA-1**$^{-2}$ and **DBA-2** species are important for the reactions A.7 and A.8 because these were performed at solution pH's higher than the respective pK_as of the reactants (see Table 3.1, and Figures A.1(b)-(c)). It is known that the product from reactions of PBA with diols are stronger acids than the PBA.[19,46] In our case, the pK_as of **DBA-1.Glc**$^{-2}$ and **DBA-2.Glc** should be lower than the pK_as of **DBA-1**$^{-2}$ and **DBA-2**, respectively. Then, we may infer that only the products present in solutions for reactions A.7 and A.8, should be the species **DBA-1.Glc**$^{-2}$ and **DBA-2.Glc**, and it is safe to ignore the corresponding equilibrium reactions A.10 and A.12.

A.2.5 Pyranose $\rightleftharpoons$ Furanose Isomerization

Bielecki and coworkers[209] demonstrated that i) glucose reacts in pyranose form with the DBA 9,10-Bis(N-methyl-N-[o-(dihydroxyboryl)benzyl]amino-methyl)anthracene, ii) the pyranose $\rightarrow$ furanose isomerization of glucose in the complex DBA.Glc occurs after the binding reaction, and iii) the reaction kinetics for this isomerization is faster in aqueous than in apolar media.[209] Therefore, the pyranose $\rightleftharpoons$ furanose equilibrium for glucose is not important for DBAs reactions and can be excluded from the experimental equilibrium constants K_{obs}. Due to the structural and chemical similarities between glucose and galactose, we consider that is also reasonable to discard the pyranose $\rightleftharpoons$ furanose isomerization as a side equilibrium reaction for galactose in K_{obs}.

A.3 Dunn's Test Comparison of CENSO's Setups

Table A.1: Dunn's test for Settings A, B, and C. The null hypothesis states that the given samples have the same median. P values were adjusted with the Bonferroni method. Level of significance: $p > 0.05$.

	A	B	C
A	-	False	False
B	False	-	False
C	False	False	-

Table A.2: Dunn's test for Settings C, E, F, and G. The null hypothesis states that the given samples have the same median. P values were adjusted with the Bonferroni method. Level of significance: $p > 0.05$.

	C	E	F	G
C	-	False	False	False
E	False	-	False	False
F	False	False	-	False
G	False	False	False	-

A.4 Additional Information for the DBA-1^{-2}+Glc reaction

Table A.3: Free energies $\Delta_r G^\circ_{\mathrm{CE}}$ for the reaction **DBA-1**$^{-2}$+Glc $\overset{K}{\rightleftharpoons}$ **DBA-1.Glc**$^{-2}$+4 H_2O with an increasing number of water molecules for microsolvation ($(H_2O)_n$). $\Delta_r G^\circ_{\mathrm{CE}}$, IQR, and $\Delta\Delta_r G^\circ_{\mathrm{CE\text{-}CD}}$ reported in kcal/mol.

	Setting B			**Setting E2**		
$(H_2O)_n$	$\Delta_r G^\circ_{\mathrm{CE}}$	IQR	$\Delta\Delta_r G^\circ_{\mathrm{CE\text{-}CD}}$	$\Delta_r G^\circ_{\mathrm{CE}}$	IQR	$\Delta\Delta_r G^\circ_{\mathrm{CE\text{-}CD}}$
0	-9.2	0.2	-3.4	-12.6	0.1	-6.8
2	-6.6	1.1	-0.8	-9.2	0.9	-3.4
3	-5.9	1.5	-0.1	-8.4	1.5	-2.6
4	-4.6	2.6	1.2	-7.3	2.7	-1.5
5	-3.1	1.3	2.8	-5.3	1.1	0.5
6	-3.7	2.1	2.1	-5.9	1.8	-0.1
7	-2.5	1.4	3.3	-4.9	2.7	0.9
10	-3.1	1.4	2.7	-4.1	1.9	1.7
13	-3.1	2.7	2.7	-3.3	2.5	2.5
16	0.4	3.3	6.2	-4.7	3.6	1.1
	Setting F2			**Setting G2**		
$(H_2O)_n$	$\Delta_r G^\circ_{\mathrm{CE}}$	IQR	$\Delta\Delta_r G^\circ_{\mathrm{CE\text{-}CD}}$	$\Delta_r G^\circ_{\mathrm{CE}}$	IQR	$\Delta\Delta_r G^\circ_{\mathrm{CE\text{-}CD}}$
0	-9.8	0.3	-4.0	-12.6	0.1	-6.8
2	-7.2	2.8	-1.4	-9.3	0.6	-3.5
3	-5.0	0.9	0.8	-7.2	1.7	-1.4
4	-2.9	1.9	2.9	-6.5	1.5	-0.7
5	-4.3	1.5	1.5	-5.2	3.6	0.6
6	-2.5	1.7	3.3	-3.7	1.0	2.1
7	-4.2	2.1	1.6	-2.3	2.0	3.5
10	-2.9	1.6	3.0	-3.1	2.1	2.7
13	-3.9	2.5	1.9	-4.1	2.1	1.7
16	-2.1	2.5	3.7	-4.9	1.7	1.0

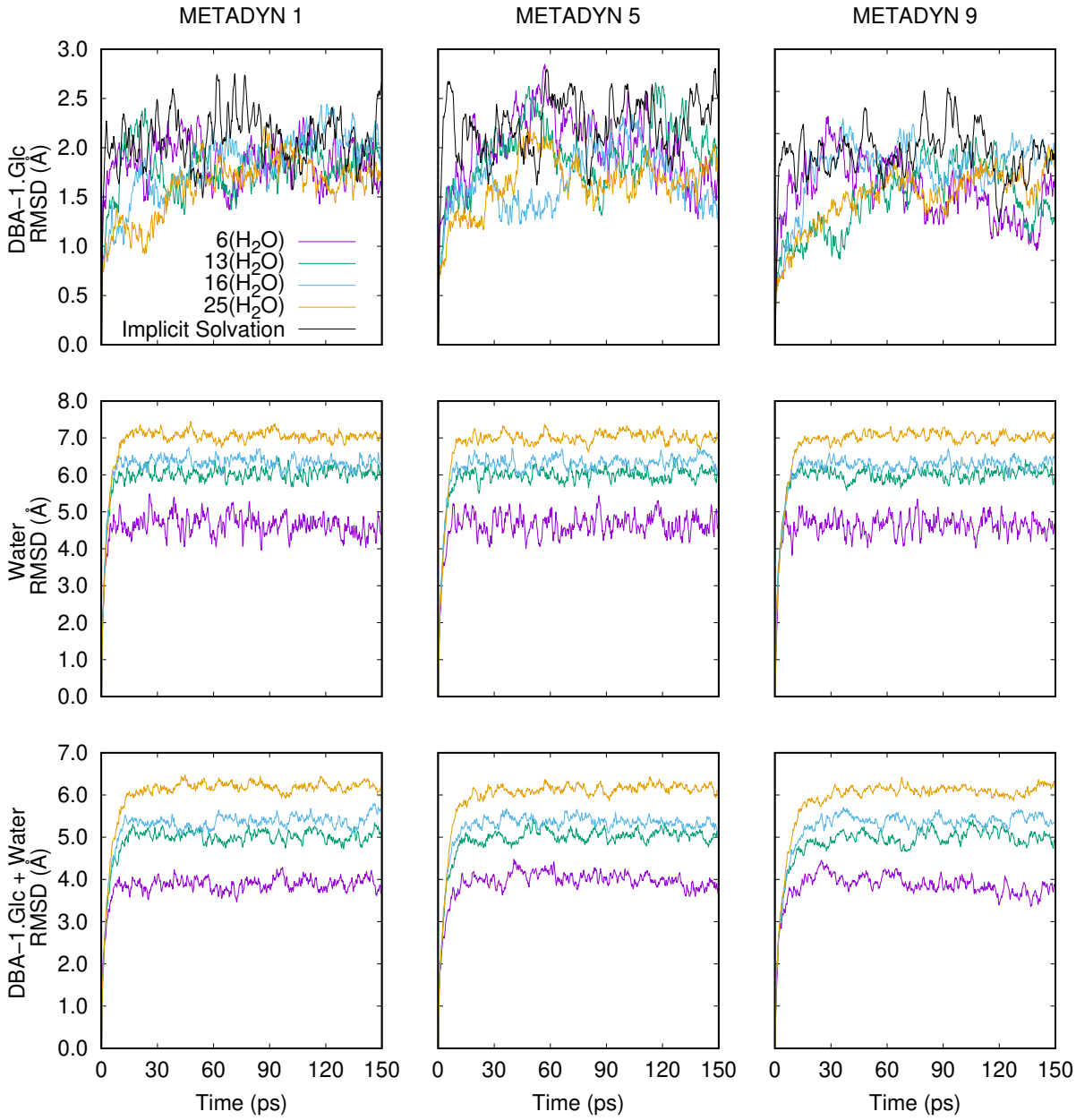


Figure A.2: RMSDs in cartesian space during NCI-MTD simulations for conformational sampling of the **DBA-1.Glc**$^{-2}$ complex. RMSDs are reported for the implicit solvation and microsolvation with 6, 13, 16, and 25 water molecules. The top row shows the RMSDs of solute **DBA-1.Glc**$^{-2}$ (for both solvation models), the middle row shows the RMSDs of the water molecules for microsolvation, and the bottom row shows the RMSDs for the complete models **DBA-1.Glc**$^{-2}(H_2O)_n$. The RMSDs were calculated from the average of ten NCI-MTD trajectories in the last iteration for conformational sampling with CREST program.[110]

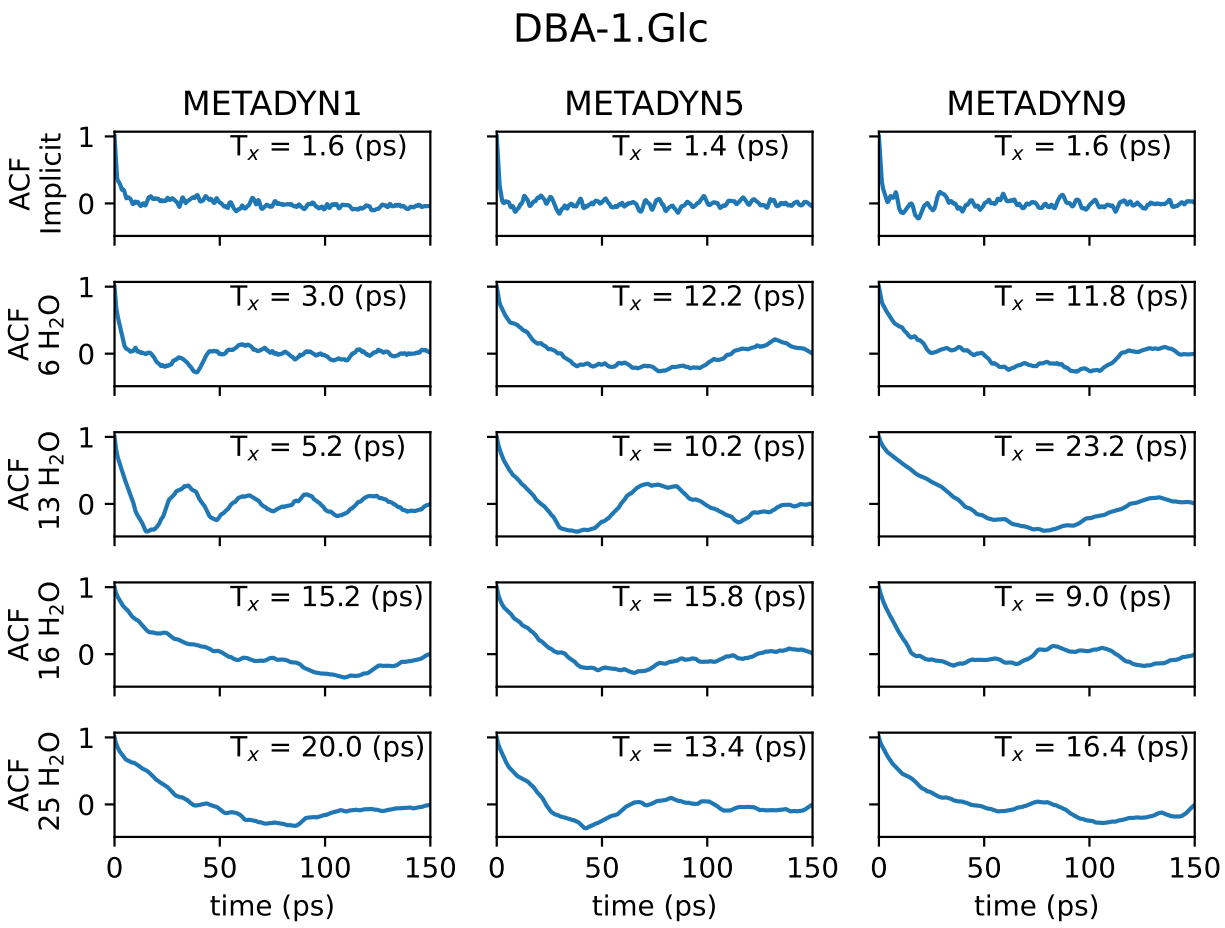


Figure A.3: ACFs for the RMSDs from the conformational sampling for complex **DBA-1.Glc**$^{-2}$ (the top row of Figure A.2) under the implicit solvent and microsolvation approximations. T_x is the corresponding time constant for the initial decay of the ACF, which is approximated by the equation: ACF(t) = $\exp(-t/T_x)$, where t is the time. T_x can be considered as the time required for no correlation (or loss of memory) from the initial RMSDs value.

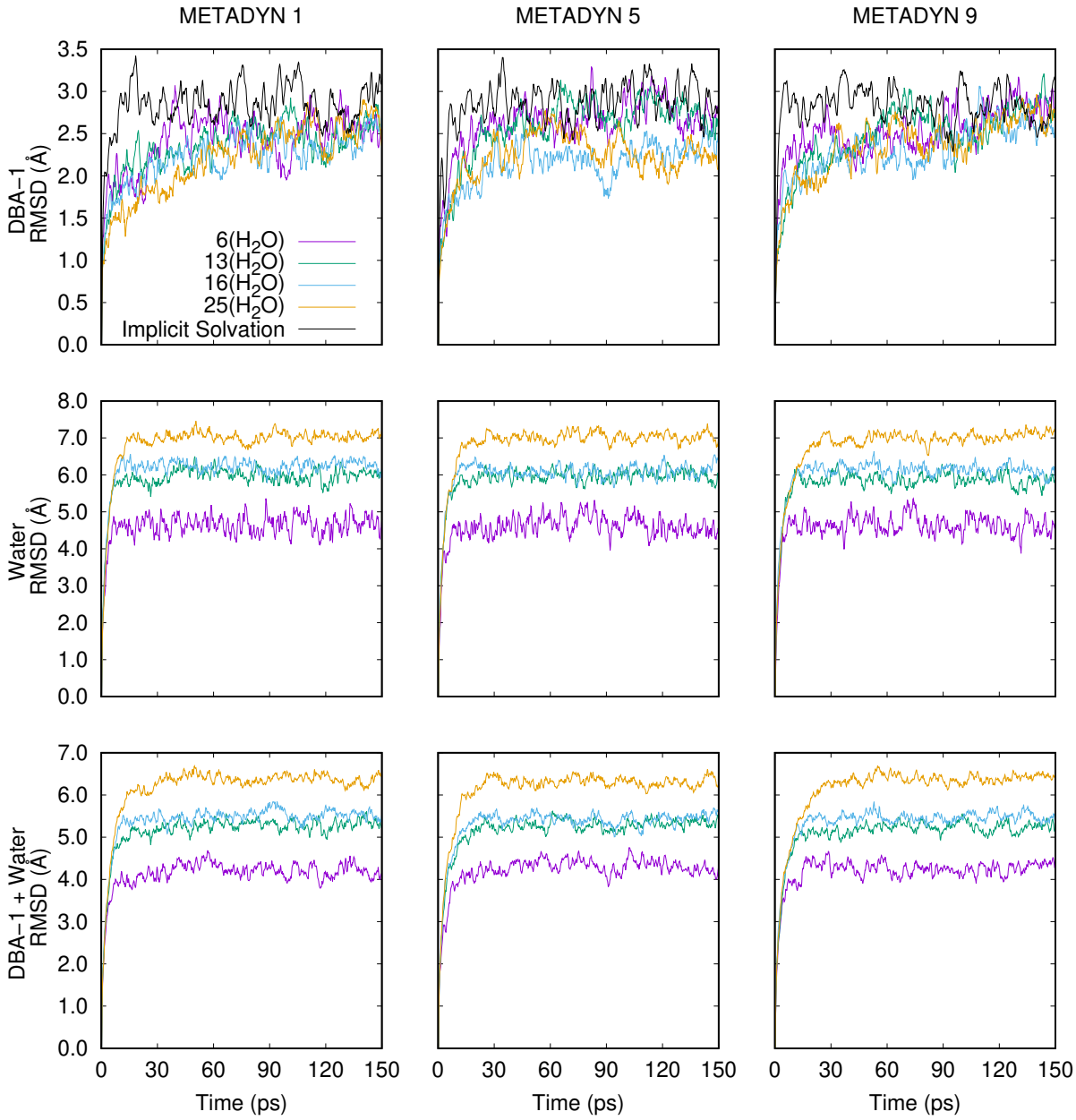


Figure A.4: RMSDs in cartesian space during NCI-MTD simulations for the conformational sampling of the free ligand **DBA-1**$^{-2}$. RMSDs are reported for the implicit solvation and microsolvation with 6, 13, 16, and 25 water molecules. The top row shows the RMSDs of solute **DBA-1**$^{-2}$,the middle row shows the RMSDs of the water molecules for microsolvation, and the bottom row shows the RMSDs for the complete models **DBA-1**$^{-2}(H_2O)_n$. RMSDs were calculated from the average of ten NCI-MTD trajectories in the last iteration for conformational sampling with the CREST program.[110]

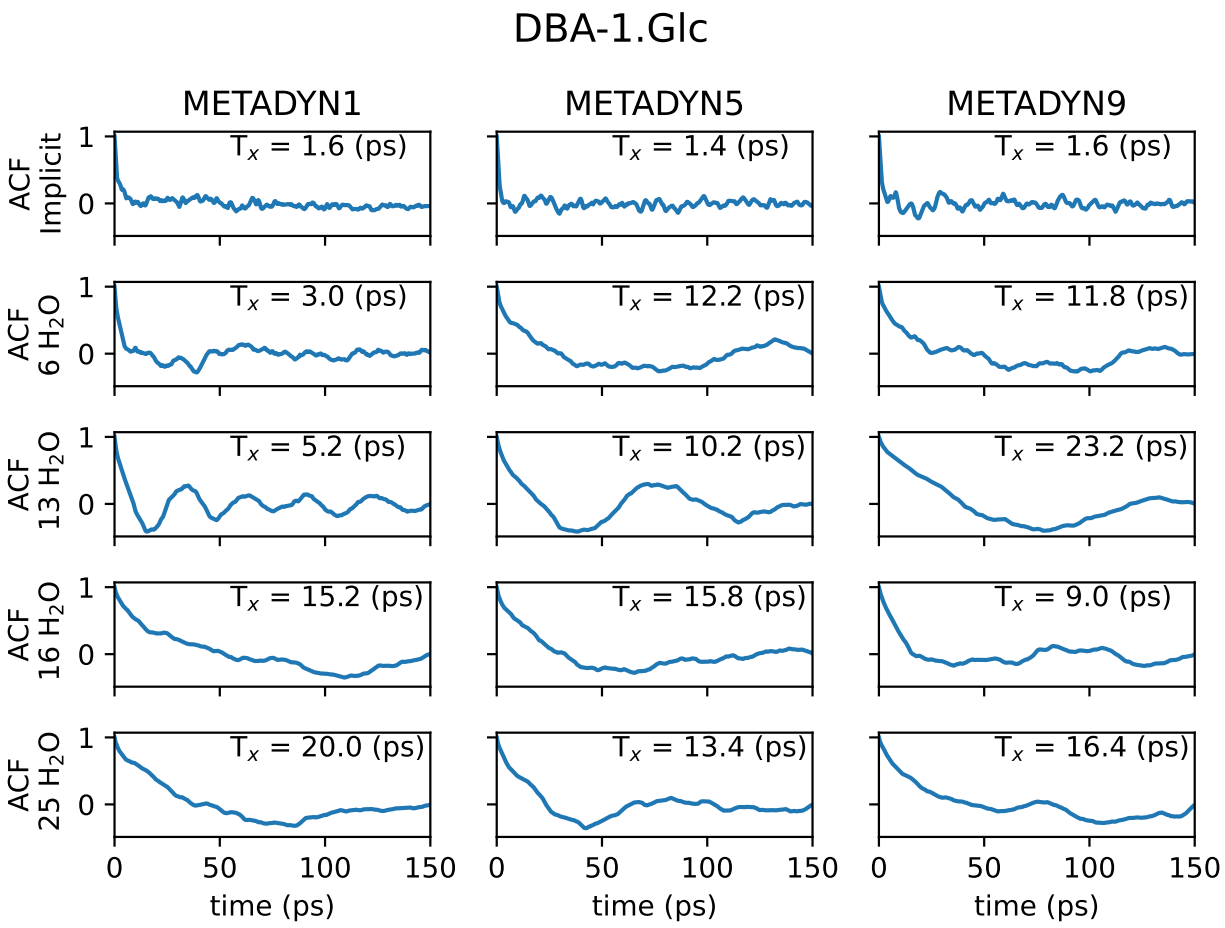


Figure A.5: ACFs for the RMSDs of **DBA-1**$^{-2}$ free ligand in top row of Figure A.4. T_x is time constant for the corresponding ACF.

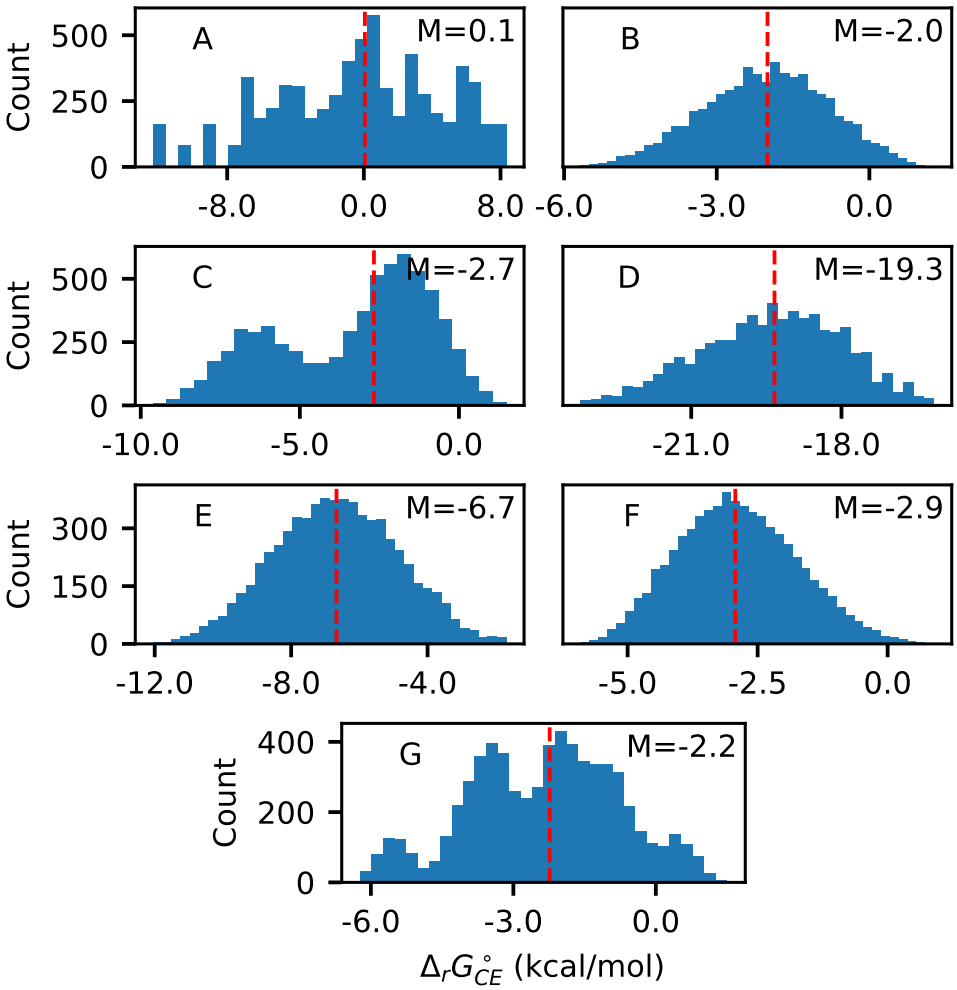


Figure A.6: Histograms distribution of the free energies $\{\Delta_r G^\circ_{\mathrm{CE},x}\}$ for reaction **DBA-1**$^{-2}$+ Glc $\overset{K}{\rightleftharpoons}$ **DBA-1.Glc**$^{-2}$ + 4 H_2O calculated with the settings reported in Table 3.2. M is the median value of $\Delta_r G^\circ_{\mathrm{CE}}$ in the corresponding histogram. The histogram's bin width was calculated with the Freedman-Diaconis rule using 6560 reaction-free energies $\Delta_r G^\circ_{\mathrm{CE},x}$ for each histogram.

Table A.4: Free energies ($G^\circ_{\mathrm{CE}}(X)$, in Hartrees) for the three CEs of **DBA-1.Glc**$^{-2}$ and **DBA-1**$^{-2}$ used for the calculation $\Delta_r G^\circ_{\mathbf{CE}}$ with the setting B, E2, F2, and G2. $(H_2O)_n$ is the number of water molecules for microsolvation. #conf is the number of conformers for the Boltzmann averages in $G^\circ(X)_{\mathrm{CE}}$. MAD and Range in kcal/mol. Average angle over the five lowest conformers $\langle \angle ABC \rangle$ in degrees.

Setting B											
$(H_2O)_n$	G°_{CE}(**DBA-1.Glc**$^{-2}$)	#conf	MAD	Range	$\langle \angle CCC \rangle$	G°_{CE}(**DBA-1**$^{-2}$)	#conf	MAD	Range	$\langle \angle CCC \rangle$	$\langle \angle BOH \rangle$
0	-1616.523268	160	0.03	0.07	110.88	-1235.207161	1857	0.01	0.02	114.08	109.56
	-1616.523275	152				-1235.207188	1885				
	-1616.52338	151				-1235.207165	1888				
2	-1769.371303	198	0.49	1.23	111.90	-1388.056538	180	0.33	0.93	113.49	109.59
	-1769.373266	82				-1388.055937	112				
	-1769.371729	214				-1388.055058	163				
3	-1845.790964	30	0.71	1.66	112.43	-1464.478682	75	0.51	1.28	112.68	109.63
	-1845.793426	64				-1464.478226	66				
	-1845.793614	67				-1464.480268	24				
4	-1922.213524	109	0.82	1.88	112.74	-1540.902836	8	1.18	2.72	113.45	109.78
	-1922.216528	14				-1540.898508	85				
	-1922.216431	93				-1540.898701	68				
5	-1998.634959	85	0.35	0.87	113.03	-1617.322253	43	0.62	1.67	112.95	109.83
	-1998.636343	46				-1617.321343	36				
	-1998.635233	83				-1617.324005	4				
6	-2075.055729	91	0.54	1.47	112.48	-1693.739726	48	1.05	2.60	112.95	109.79
	-2075.058069	40				-1693.740451	62				
	-2075.057235	50				-1693.743863	24				
7	-2151.478554	74	0.33	0.82	112.51	-1770.161515	75	0.09	0.20	112.1	109.91
	-2151.47831	21				-1770.161811	47				
	-2151.479613	5				-1770.161493	23				
10	-2380.739023	2	0.53	1.24	112.80	-1999.425003	14	0.64	1.54	110.79	110.11
	-2380.740827	10				-1999.422899	24				
	-2380.740996	8				-1999.422549	17				
13	-2609.999609	10	0.78	1.85	113.90	-2228.684696	6	1.11	2.62	112.53	110.28
	-2610.002228	21				-2228.688868	9				
	-2610.002559	10				-2228.685094	14				
16	-2839.260914	12	0.82	1.95	113.37	-2457.952151	2	1.61	4.64	112.17	110.47
	-2839.264015	2				-2457.948002	4				
	-2839.261215	10				-2457.944759	5				

Continuation of Table A.4

Setting E2											
$(H_2O)_n$	G°_{CE}(**DBA-1.Glc**$^{-2}$)	#conf	MAD	Range	$\langle\measuredangle CCC\rangle$	G°_{CE}(**DBA-1**$^{-2}$)	#conf	MAD	Range	$\langle\measuredangle CCC\rangle$	$\langle\measuredangle BOH\rangle$
	-1616.3697652	167				-1235.1135629	2262				
0	-1616.3697645	166	0.04	0.08	112.72	-1235.1134834	2326	2.15×10^{-2}	4.99×10^{-2}	113.39	108.45
	-1616.3698920	169				-1235.1135584	2282				
	-1769.2316842	366				-1387.9761847	280				
2	-1769.2329356	351	0.36	0.85	112.03	-1387.9761358	162	0.39	0.88	113.43	108.50
	-1769.2330456	307				-1387.9775450	204				
	-1845.6606767	114				-1464.4045139	56				
3	-1845.6605788	125	0.16	0.40	112.42	-1464.4051236	71	0.52	1.36	112.75	108.49
	-1845.6612107	128				-1464.4066808	30				
	-1922.0877707	33				-1540.8344399	141				
4	-1922.0873155	149	0.29	0.79	113.13	-1540.8313258	20	1.12	3.11	112.41	108.79
	-1922.0885751	59				-1540.8294907	46				
	-1998.5151547	52				-1617.2603408	5				
5	-1998.5143823	46	0.26	0.70	113.25	-1617.2615905	22	0.55	1.62	113.50	108.76
	-1998.5155027	32				-1617.2590090	33				
	-2074.9409472	57				-1693.6838932	6				
6	-2074.9410600	20	0.65	1.50	113.15	-1693.6846911	13	0.41	1.17	111.77	108.85
	-2074.9433349	27				-1693.6857541	11				
	-2151.3681437	40				-1770.1154924	17				
7	-2151.3679282	30	1.24	2.85	113.35	-1770.1103720	27	1.08	3.21	112.48	108.89
	-2151.3724654	24				-1770.1129881	15				
	-2380.6462929	20				-1999.3903506	23				
10	-2380.6477712	10	0.62	1.85	112.99	-1999.3937283	7	0.93	2.12	111.54	109.05
	-2380.6492393	15				-1999.3904225	4				
	-2609.9278439	8				-2228.6772469	15				
13	-2609.9281134	5	1.00	2.33	113.11	-2228.6754064	5	1.28	3.46	113.07	109.59
	-2609.9315628	14				-2228.6717400	8				
	-2839.2062350	5				-2457.9468929	1				
16	-2839.2047299	1	1.10	2.95	113.85	-2457.9476570	4	0.81	2.05	113.20	109.64
	-2839.2094383	3				-2457.9501666	4				

Continuation of Table A.4

Setting F2											
$(H_2O)_n$	G°_{CE}(**DBA-1.Glc**$^{-2}$)	#conf	MAD	Range	$\langle \angle CCC \rangle$	G°_{CE}(**DBA-1**$^{-2}$)	#conf	MAD	Range	$\langle \angle CCC \rangle$	$\langle \angle BOH \rangle$
0	-1616.513756	121	0.14	0.42	112.21	-1235.202342	1765	3.71×10^{-3}	0.01	113.14	111.84
	-1616.514426	116				-1235.202331	1814				
	-1616.514111	117				-1235.202348	1771				
2	-1769.363335	201	1.31	3.12	111.65	-1388.049187	144	0.53	1.40	111.31	112.38
	-1769.358892	128				-1388.049872	127				
	-1769.363864	92				-1388.051415	154				
3	-1845.779985	80	0.46	1.35	112.34	-1464.470015	236	0.60	1.42	111.77	112.45
	-1845.778854	126				-1464.470255	108				
	-1845.777842	158				-1464.472283	54				
4	-1922.204435	63	0.92	2.51	112.27	-1540.895261	6	0.70	2.03	111.20	112.51
	-1922.200429	113				-1540.89345	45				
	-1922.201826	40				-1540.892022	28				
5	-1998.623712	21	0.83	2.22	111.47	-1617.314395	23	0.52	1.26	110.71	112.62
	-1998.622565	75				-1617.312695	36				
	-1998.626108	12				-1617.312393	29				
6	-2075.046199	32	0.48	1.27	112.20	-1693.737075	8	0.82	1.88	111.77	112.71
	-2075.044783	18				-1693.734075	18				
	-2075.044177	31				-1693.736926	16				
7	-2151.465172	48	0.35	1.00	112.15	-1770.156605	5	1.00	2.56	109.76	112.80
	-2151.465866	16				-1770.153532	23				
	-2151.46427	39				-1770.152518	19				
10	-2380.725169	2	0.10	0.27	111.41	-1999.415668	8	0.87	2.16	110.13	113.09
	-2380.725597	4				-1999.412228	8				
	-2380.725455	7				-1999.412856	2				
13	-2609.986078	5	1.31	3.45	111.63	-2228.673123	7	0.08	0.24	111.33	113.05
	-2609.987678	4				-2228.673321	3				
	-2609.982174	3				-2228.673502	3				
16	-2839.243291	2	0.65	1.92	112.01	-2457.934803	4	0.53	1.36	112.45	113.10
	-2839.244719	2				-2457.933163	3				
	-2839.241666	3				-2457.932633	2				

Continuation of Table A.4

Setting G2											
$(H_2O)_n$	G°_{CE}(**DBA-1.Glc**$^{-2}$)	#conf	MAD	Range	$\langle\measuredangle CCC\rangle$	G°_{CE}(**DBA-1**$^{-2}$)	#conf	MAD	Range	$\langle\measuredangle CCC\rangle$	$\langle\measuredangle BOH\rangle$
	-1616.3697541	168				-1235.1135608	2275				
0	-1616.3696993	163	0.01	0.03	112.65	-1235.1135591	2355	6.69×10^{-4}	1.95×10^{-3}	113.30	108.44
	-1616.3697234	162				-1235.1135577	2294				
	-1769.2297338	206				-1387.9730277	226				
2	-1769.2312956	244	0.41	0.98	111.16	-1387.9749572	180	0.45	1.21	113.53	108.52
	-1769.2299295	150				-1387.9736408	133				
	-1845.6567998	42				-1464.4027535	47				
3	-1845.6571108	78	0.40	1.00	112.32	-1464.4004955	51	0.84	2.37	113.74	108.54
	-1845.6583933	38				-1464.4042672	46				
	-1922.0863437	40				-1540.8283575	31				
4	-1922.0848314	79	0.38	0.95	113.23	-1540.8307103	89	0.54	1.48	113.08	108.64
	-1922.0860120	58				-1540.8298458	98				
	-1998.5129264	65				-1617.2591341	25				
5	-1998.5125381	48	0.33	0.86	112.59	-1617.2581519	35	0.40	1.18	113.34	108.70
	-1998.5115513	61				-1617.2572534	29				
	-2074.9396763	79				-1693.6870610	26				
6	-2074.9391613	43	0.36	0.97	113.27	-1693.6852972	38	0.41	1.11	112.21	108.92
	-2074.9407140	41				-1693.6864683	45				
	-2151.3665298	41				-1770.1116786	16				
7	-2151.3657242	51	0.20	0.51	112.88	-1770.1116398	38	0.52	1.18	113.36	108.92
	-2151.3658815	15				-1770.1135237	26				
	-2380.6469975	11				-1999.3889605	16				
10	-2380.6488026	29	0.74	2.21	112.99	-1999.3914843	31	0.93	2.62	112.59	109.15
	-2380.6452829	11				-1999.3931357	15				
	-2609.9252401	11				-2228.6668632	1				
13	-2609.9245847	19	0.24	0.67	112.98	-2228.6689241	1	0.53	1.29	112.05	109.37
	-2609.9256454	6				-2228.6686323	4				
	-2839.2056008	4				-2457.9475955	3				
16	-2839.2085998	11	0.87	2.04	112.78	-2457.9479901	4	0.16	0.47	111.59	109.74
	-2839.2053488	4				-2457.9483465	1				

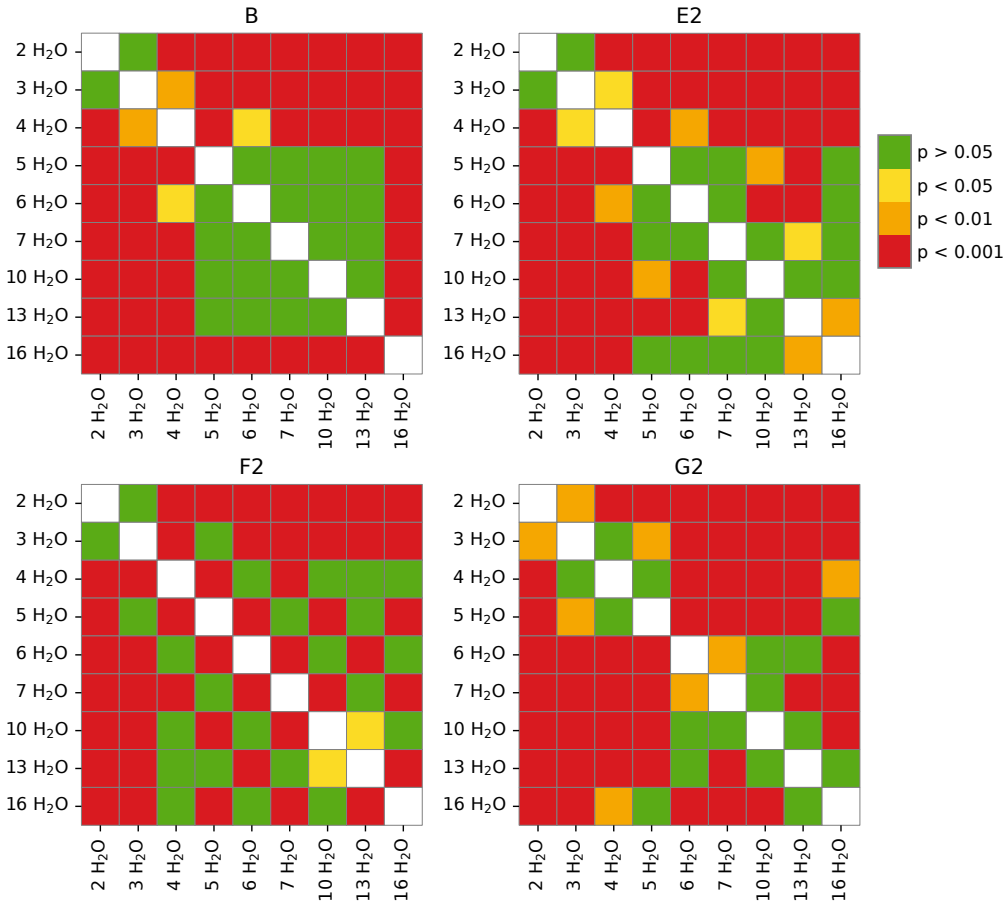


Figure A.7: p-values from Dunn's test for multi-pairwise comparisons for the distributions $\{\Delta_r G^{\circ}_{\mathrm{CE},x}\}$ for the reaction $\mathbf{DBA\text{-}1}^{-2} + \mathrm{Glc} \overset{K}{\rightleftharpoons} \mathbf{DBA\text{-}1.Glc}^{-2} + 4\,\mathrm{H_2O}$, and reported Figure 4.2. The comparisons are performed across setups B, E2, F2, and G2.

A.5 Additional Information for the DBA-2+Glc Reaction

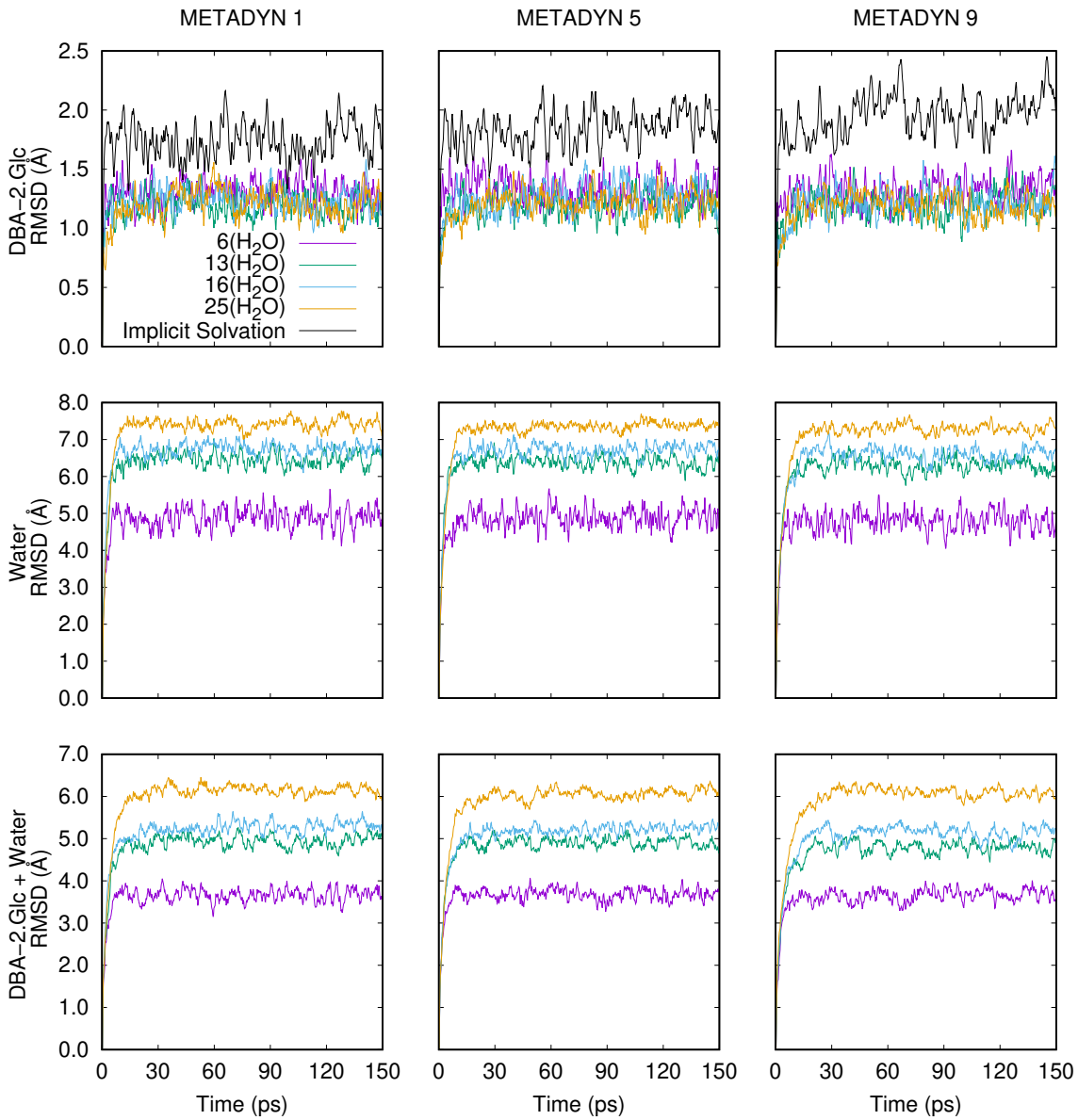


Figure A.8: RMSDs in cartesian space during NCI-MTD simulations for the conformational sampling of the complex **DBA-2.Glc**. RMSDs are reported for the implicit solvation and microsolvation with 6, 13, 16, and 25 water molecules. The top row shows the RMSDs of solute **DBA-2.Glc**, the middle row shows the RMSDs of the water molecules for microsolvation, and the bottom row shows the RMSDs for the complete models **DBA-2.Glc**$(H_2O)_n$. RMSDs were calculated from the average of ten NCI-MTD trajectories in the last iteration for conformational sampling with the CREST program.[110]

Table A.5: Free energies $\Delta_r G^\circ_{\text{CE}}$ for the reaction **DBA-2** + Glc $\overset{K}{\rightleftharpoons}$ **DBA-2.Glc** + 4 H_2O with an increasing number of water molecules for microsolvation ($(H_2O)_n$). $\Delta_r G^\circ_{\text{CE}}$, IQR, $\Delta\Delta_r G^\circ_{\text{CE-PT}}$, and $\Delta\Delta_r G^\circ_{\text{CE-NMR}}$ reported in kcal/mol.

Setting B					**Setting E2**			
$(H_2O)_n$	$\Delta_r G^\circ_{\text{CE}}$	IQR	$\Delta\Delta_r G^\circ_{\text{CE-PT}}$	$\Delta\Delta_r G^\circ_{\text{CE-NMR}}$	$\Delta_r G^\circ_{\text{CE}}$	IQR	$\Delta\Delta_r G^\circ_{\text{CE-PT}}$	$\Delta\Delta_r G^\circ_{\text{CE-NMR}}$
0	-13.3	0.3	-8.7	-7.8	-16.8	0.2	-12.2	-11.3
2	-8.4	0.7	-3.8	-2.9	-13.3	1.3	-8.7	-7.8
3	-9.8	0.7	-5.2	-4.3	-12.5	3.1	-7.9	-7.0
4	-8.2	1.1	-3.6	-2.7	-8.5	3.7	-3.9	-3.0
5	-8.8	1.9	-4.2	-3.3	-10.9	3.1	-6.3	-5.4
6	-8.4	1.2	-3.8	-2.9	-9.5	2.1	-4.9	-4.0
7	-6.8	2.3	-2.2	-1.3	-10.0	1.1	-5.4	-4.5
10	-5.2	2.3	-0.6	0.3	-8.3	1.8	-3.7	-2.8
13	-4.0	3.1	0.6	1.5	-4.9	2.0	-0.3	0.6
16	-1.2	2.6	3.4	4.3	-5.3	3.4	-0.7	0.2
Setting F2					**Setting G2**			
$(H_2O)_n$	$\Delta_r G^\circ_{\text{CE}}$	IQR	$\Delta\Delta_r G^\circ_{\text{CE-PT}}$	$\Delta\Delta_r G^\circ_{\text{CE-NMR}}$	$\Delta_r G^\circ_{\text{CE}}$	IQR	$\Delta\Delta_r G^\circ_{\text{CE-PT}}$	$\Delta\Delta_r G^\circ_{\text{CE-NMR}}$
0	-14.4	0.2	-9.8	-8.9	-17.0	0.2	-12.4	-11.5
2	-7.9	1.6	-3.3	-2.4	-13.1	2.2	-8.5	-7.6
3	-9.3	1.4	-4.7	-3.8	-11.2	1.3	-6.6	-5.7
4	-7.4	2.1	-2.8	-1.9	-9.8	2.3	-5.2	-4.3
5	-8.8	1.8	-4.2	-3.3	-9.5	3.5	-4.9	-4.0
6	-8.6	2.4	-4.0	-3.1	-9.1	1.4	-4.5	-3.6
7	-10.2	4.2	-5.6	-4.7	-6.8	2.5	-2.2	-1.3
10	-5.3	2.1	-0.7	0.3	-7.3	1.2	-2.7	-1.8
13	-4.9	2.9	-0.3	0.6	-5.9	4.2	-1.3	-0.4
16	-5.6	3.6	-1.0	-0.1	-7.3	4.6	-2.7	-1.8

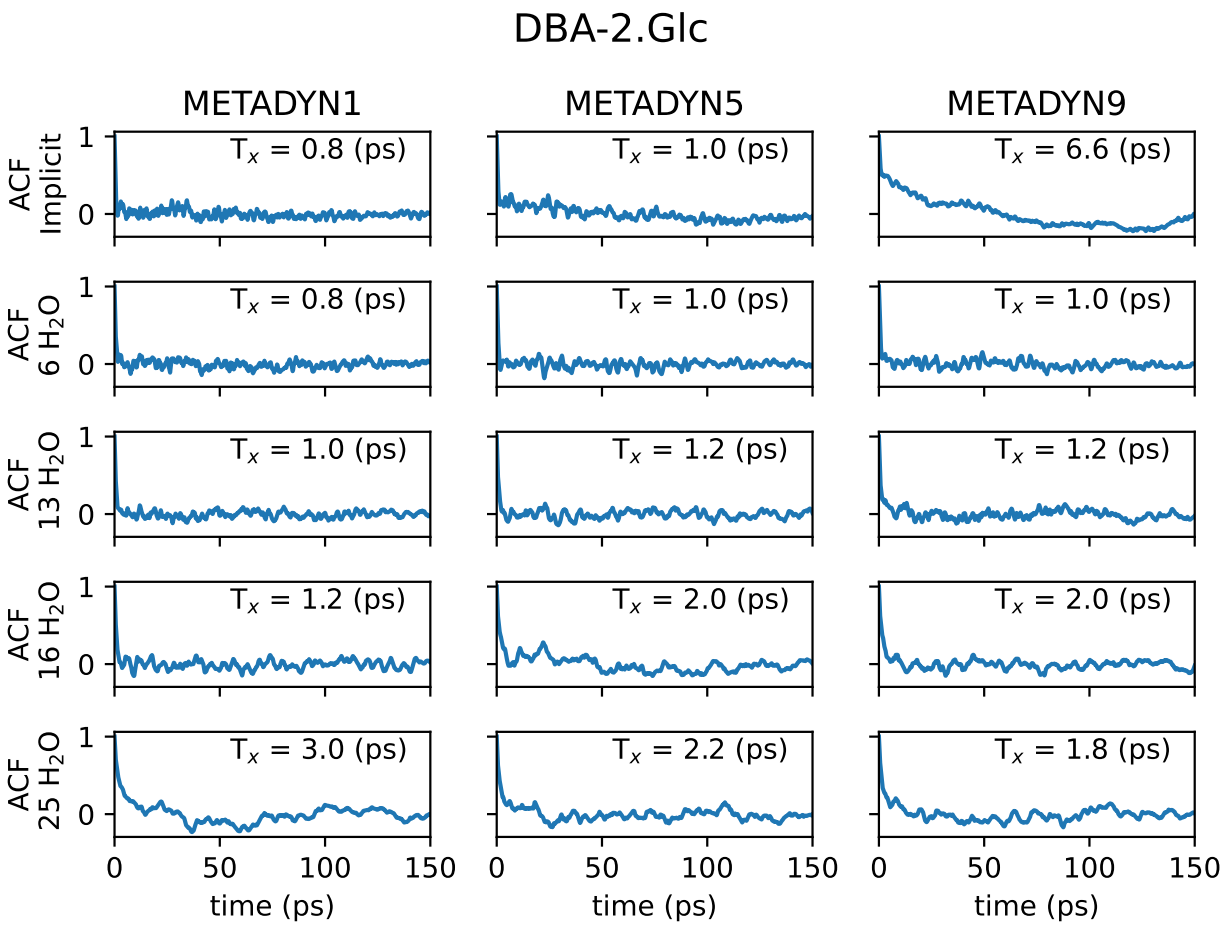


Figure A.9: ACFs for the RMSDs of DBA-2.Glc complex in top row of Figure A.8. T_x is time constant for the corresponding ACF.

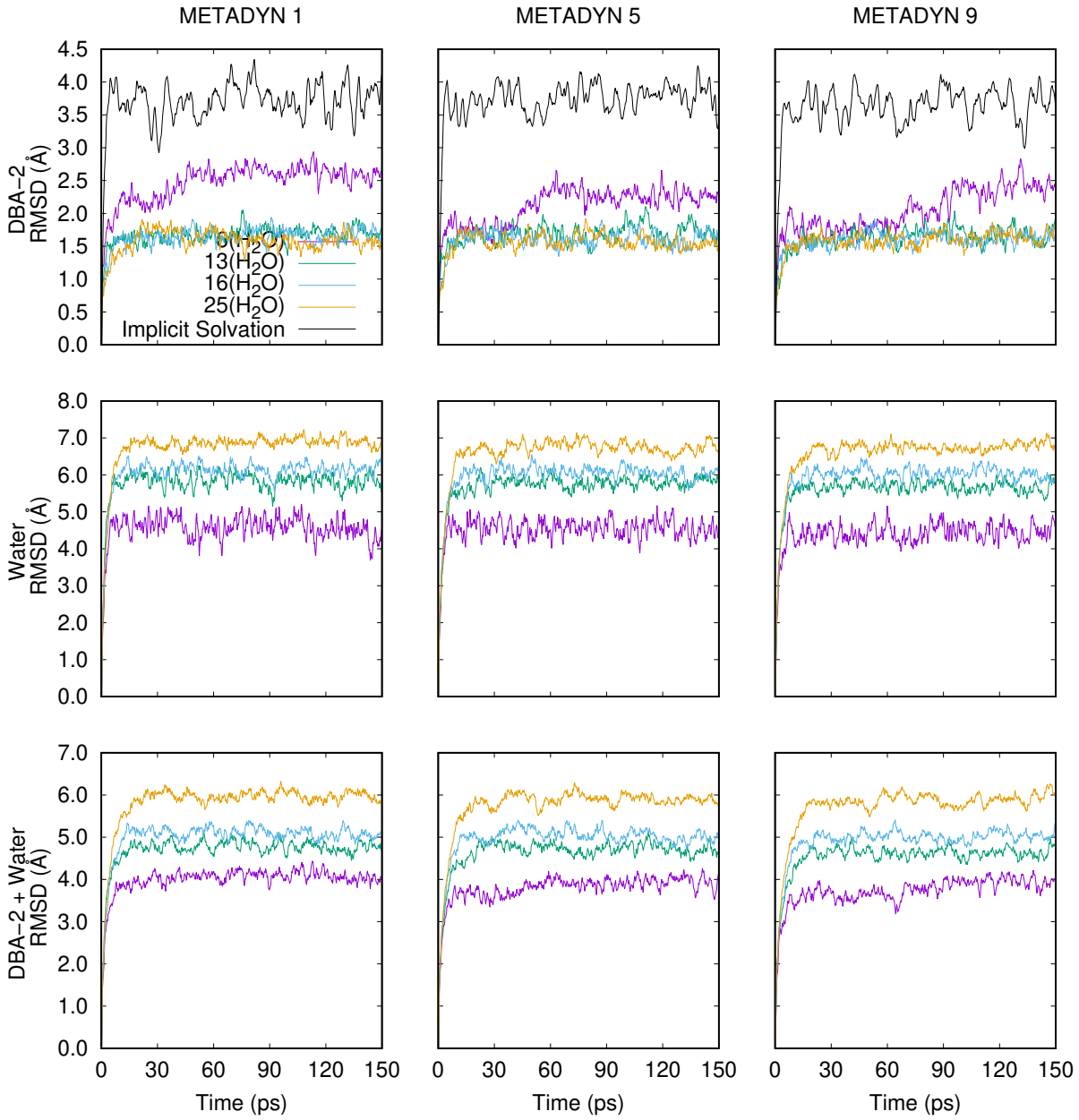


Figure A.10: RMSDs in cartesian space during NCI-MTD simulations for the conformational sampling of the free ligand **DBA-2**. RMSDs are reported for the implicit solvation and microsolvation with 6, 13, 16, and 25 water molecules. The top row shows the RMSDs of solute **DBA-2**, the middle row shows the RMSDs of the water molecules for microsolvation, and the bottom row shows the RMSDs for the complete models **DBA-2**$(H_2O)_n$. RMSDs were calculated from the average of ten NCI-MTD trajectories in the last iteration for conformational sampling with the CREST program.[110]

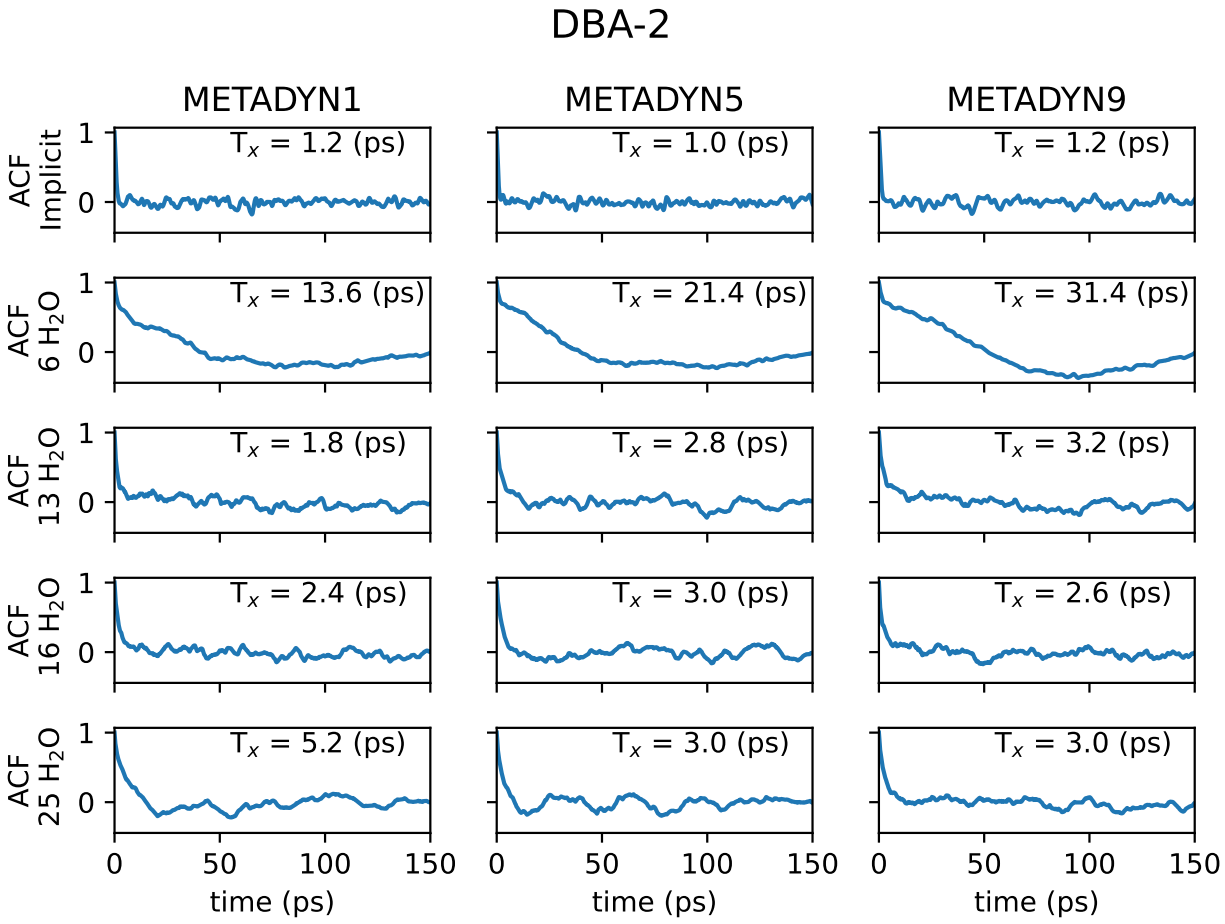


Figure A.11: ACFs for the RMSDs of DBA-2 free ligand in top row of Figure A.10. T_x is time constant for the corresponding ACF.

Table A.6: Free energies ($G^{\circ}_{\mathrm{CE}}(X)$), in Hartrees) for the three CEs of **DBA-2.Glc** and **DBA-2** used for the calculation $\Delta_r G^{\circ}_{\mathrm{CE}}$ with setting B, E2, F2, and G2. $(H_2O)_n$ is the number of water molecules for microsolvation. #conf is the number of conformers for the Boltzmann average in $G^{\circ}_{\mathrm{CE}}(X)$. MAD and Range in kcal/mol. Average angle over the five lowest conformers $\langle \measuredangle ABC \rangle$ in degrees.

Setting B											
$(H_2O)_n$	G°_{CE}(**DBA-2.Glc**)	#conf	MAD	Range	$\langle \measuredangle NCC \rangle$	G°_{CE}(**DBA-2**)	#conf	MAD	Range	$\langle \measuredangle NCC \rangle$	$\langle \measuredangle BOH \rangle$
	-1997.88721	1857				-1616.566864	2262				
0	-1997.88746	1885	0.07	0.16	112.36	-1616.566977	2326	0.10	0.27	111.82	109.97
	-1997.887236	1888				-1616.567294	2282				
	-2150.73631	180				-1769.419259	280				
2	-2150.736125	112	0.35	0.84	112.7	-1769.419416	162	0.04	0.10	111.13	110.14
	-2150.734965	163				-1769.419362	204				
	-2227.156542	75				-1845.840627	56				
3	-2227.157931	66	0.35	0.87	112.76	-1845.840211	71	0.16	0.46	112.87	110.16
	-2227.157628	24				-1845.840951	30				
	-2303.580577	8				-1922.260422	141				
4	-2303.577912	85	0.74	1.67	112.78	-1922.260408	20	0.11	0.25	112.61	110.35
	-2303.577967	68				-1922.260022	46				
	-2380.000134	43				-1998.680991	5				
5	-2380.000348	36	0.39	0.94	112.84	-1998.68408	22	0.95	2.32	112.87	110.34
	-2380.001626	4				-1998.680376	33				
	-2456.423169	48				-2075.100708	6				
6	-2456.419944	62	0.70	2.02	112.84	-2075.100988	13	0.07	0.18	112.96	110.57
	-2456.421384	24				-2075.100781	11				
	-2532.842196	75				-2151.521116	17				
7	-2532.842571	47	1.09	2.58	112.78	-2151.522724	27	0.57	1.57	113.34	110.59
	-2532.846307	23				-2151.523617	15				
	-2762.102129	14				-2380.783947	23				
10	-2762.103155	24	1.19	3.01	112.93	-2380.784746	7	0.56	1.50	113.26	110.84
	-2762.098363	17				-2380.782349	4				
	-2991.357272	6				-2610.04448	15				
13	-2991.359964	9	0.97	2.67	113.18	-2610.043934	5	1.30	3.09	113.61	111.12
	-2991.361521	14				-2610.048861	8				
	-3220.615097	2				-2839.303777	1				
16	-3220.616962	4	0.75	2.21	113.18	-2839.304596	4	0.84	2.16	113.08	111.07
	-3220.618618	5				-2839.301158	4				

Continuation of Table A.6.

Setting E2											
$(H_2O)_n$	G°_{CE}(**DBA-2.Glc**)	#conf	MAD	Range	$\langle \measuredangle NCC \rangle$	G°_{CE}(**DBA-2**)	#conf	MAD	Range	$\langle \measuredangle NCC \rangle$	$\langle \measuredangle BOH \rangle$
	-1997.694413	12				-1616.433636	120				
0	-1997.694192	15	0.06	0.14	112.32	-1616.433709	133	0.06	0.17	111.05	108.93
	-1997.694234	15				-1616.433443	105				
	-2150.557692	88				-1769.297976	39				
2	-2150.559039	78	0.47	1.25	112.27	-1769.297966	12	0.59	1.32	111.15	109.93
	-2150.559685	91				-1769.300071	10				
	-2226.986445	20				-1845.725511	29				
3	-2226.980812	30	1.46	3.53	112.25	-1845.723138	6	1.06	3.13	112.40	109.19
	-2226.985654	58				-1845.728128	7				
	-2303.411441	16				-1922.1511	8				
4	-2303.407156	2	1.03	2.69	113.14	-1922.157459	3	1.41	3.99	113.07	109.39
	-2303.410252	15				-1922.153682	5				
	-2379.841354	10				-1998.579294	8				
5	-2379.844325	2	1.46	4.22	112.80	-1998.58039	10	0.58	1.64	112.96	106.66
	-2379.837597	7				-1998.581908	3				
	-2456.268029	7				-2075.006211	7				
6	-2456.266687	7	0.78	2.19	113.02	-2075.008628	1	0.57	1.52	112.56	109.72
	-2456.270173	4				-2075.007901	3				
	-2532.694787	1				-2151.432738	4				
7	-2532.695747	2	0.31	0.81	113.24	-2151.431764	9	0.55	1.55	112.62	109.91
	-2532.694462	6				-2151.434237	5				
	-2761.971895	3				-2380.711368	3				
10	-2761.972412	2	0.84	2.06	113.38	-2380.709842	2	0.45	1.06	112.74	110.24
	-2761.969135	3				-2380.71153	2				
	-2991.242931	2				-2609.991584	1				
13	-2991.246747	3	1.08	2.49	113.27	-2609.989216	1	0.78	2.01	113.25	110.15
	-2991.246894	3				-2609.992413	1				
	-3220.520293	4				-2839.265277	1				
16	-3220.521708	3	0.51	1.39	113.50	-2839.264032	1	0.49	1.44	113.43	110.94
	-3220.522513	1				-2839.26298	4				

Continuation of Table A.6.

	Setting F2										
$(H_2O)_n$	G°_{CE}(**DBA-2.Glc**)	#conf	MAD	Range	$\langle \angle NCC \rangle$	G°_{CE}(**DBA-2**)	#conf	MAD	Range	$\langle \angle NCC \rangle$	$\langle \angle BOH \rangle$
0	-1997.876443	121	0.08	0.20	111.26	-1616.559712	1765	0.02	0.06	109.85	112.75
	-1997.876212	116				-1616.559802	1814				
	-1997.876128	117				-1616.559786	1771				
2	-2150.724065	201	0.83	1.91	111.63	-1769.410917	144	0.43	1.22	109.41	113.13
	-2150.721016	128				-1769.409776	127				
	-2150.721168	92				-1769.411714	154				
3	-2227.144458	80	0.99	2.52	111.77	-1845.831399	236	0.68	1.72	110.85	113.33
	-2227.147538	126				-1845.82929	108				
	-2227.143518	158				-1845.828665	54				
4	-2303.565163	63	0.38	0.88	111.50	-1922.254017	6	0.87	2.57	111.02	113.35
	-2303.565081	113				-1922.249924	45				
	-2303.566486	40				-1922.252085	28				
5	-2379.989027	21	0.92	2.39	111.88	-1998.672325	23	0.38	1.08	110.94	113.46
	-2379.985214	75				-1998.673007	36				
	-2379.987983	12				-1998.671288	29				
6	-2456.406146	32	1.34	3.15	111.96	-2075.09034	8	0.59	1.45	110.76	113.46
	-2456.406596	18				-2075.092648	18				
	-2456.411166	31				-2075.090757	16				
7	-2532.832937	48	1.88	4.60	111.77	-2151.50703	5	1.29	3.40	110.56	113.76
	-2532.825602	16				-2151.512445	23				
	-2532.826756	39				-2151.510857	19				
10	-2762.080125	2	1.01	2.52	112.33	-2380.769745	8	0.77	1.88	110.67	113.96
	-2762.083333	4				-2380.767249	8				
	-2762.084143	7				-2380.770244	2				
13	-2991.338781	5	0.68	1.88	112.37	-2610.030482	7	1.27	2.99	111.43	114.13
	-2991.341777	4				-2610.025722	3				
	-2991.34068	3				-2610.030055	3				
16	-3220.59866	2	1.47	3.69	112.40	-2839.285973	4	1.34	3.36	111.63	113.92
	-3220.59278	2				-2839.284842	3				
	-3220.597458	3				-2839.280615	2				

Continuation of Table A.6.

Setting G2											
$(H_2O)_n$	G°_{CE}(**DBA-2.Glc**)	#conf	MAD	Range	$\langle\measuredangle NCC\rangle$	G°_{CE}(**DBA-2**)	#conf	MAD	Range	$\langle\measuredangle NCC\rangle$	$\langle\measuredangle BOH\rangle$
	-1997.694207	167				-1616.433352	2262				
0	-1997.694356	166	0.06	0.18	112.31	-1616.43326	2326	0.12	0.31	111.35	108.89
	-1997.694486	169				-1616.43375	2282				
	-2150.551956	366				-1769.295476	280				
2	-2150.556132	351	1.07	2.62	112.30	-1769.294756	162	0.19	0.45	111.61	109.35
	-2150.555424	307				-1769.295377	204				
	-2226.980598	114				-1845.724628	56				
3	-2226.98307	125	0.66	1.55	112.74	-1845.723093	71	0.38	0.96	111.91	109.37
	-2226.982875	128				-1845.723452	30				
	-2303.40992	33				-1922.14935	141				
4	-2303.411334	149	0.92	2.52	112.96	-1922.151047	20	0.52	1.26	112.77	109.55
	-2303.407325	59				-1922.151355	46				
	-2379.835627	52				-1998.579892	5				
5	-2379.834918	46	0.17	0.45	113.00	-1998.575228	22	1.11	2.93	113.53	109.81
	-2379.835115	32				-1998.578554	33				
	-2456.264783	57				-2075.003889	6				
6	-2456.264479	20	0.44	1.09	113.07	-2075.005884	13	0.58	1.36	112.63	109.90
	-2456.263051	27				-2075.003721	11				
	-2532.691355	40				-2151.432429	17				
7	-2532.692073	30	0.76	1.94	112.93	-2151.429401	27	1.11	3.08	113.32	109.57
	-2532.688981	24				-2151.434303	15				
	-2761.966317	20				-2380.705812	23				
10	-2761.966657	10	0.12	0.33	113.18	-2380.707609	7	0.49	1.13	112.19	110.10
	-2761.96685	15				-2380.705882	4				
	-2991.239849	8				-2609.986846	15				
13	-2991.240848	5	0.25	0.63	113.35	-2609.982925	5	2.16	6.09	113.14	110.33
	-2991.240646	14				-2609.977139	8				
	-3220.51951	5				-2839.257089	1				
16	-3220.516198	1	1.60	4.63	113.78	-2839.266152	4	2.55	5.79	112.44	110.49
	-3220.523575	3				-2839.256926	4				

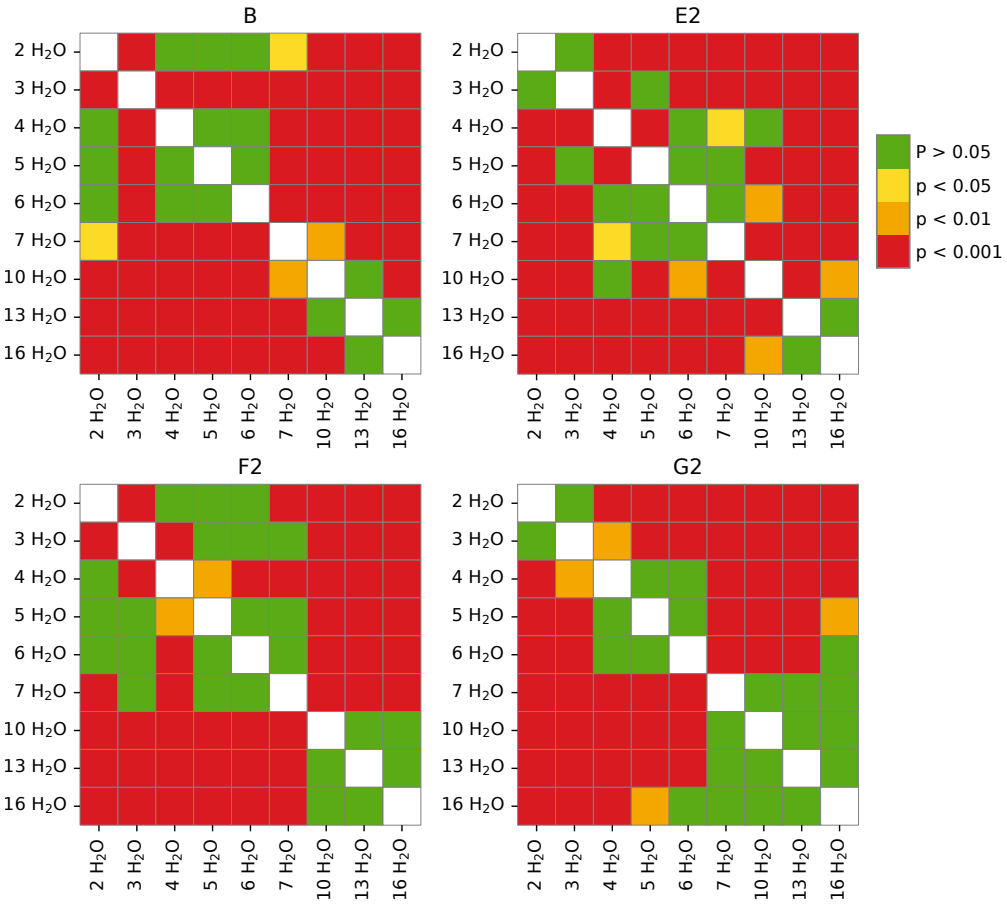


Figure A.12: p-values from Dunn's test for multi-pairwise comparisons for the distributions $\{\Delta_r G^{\circ}_{\mathrm{CE},x}\}$ for the reaction **DBA-2** + Glc $\overset{K}{\rightleftharpoons}$ **DBA-2.Glc** + 4 H_2O, and reported Figure 4.3. The comparisons are performed across setups B, E2, F2, and G2.

A.6 Additional Information for Glucose and Water Clusters

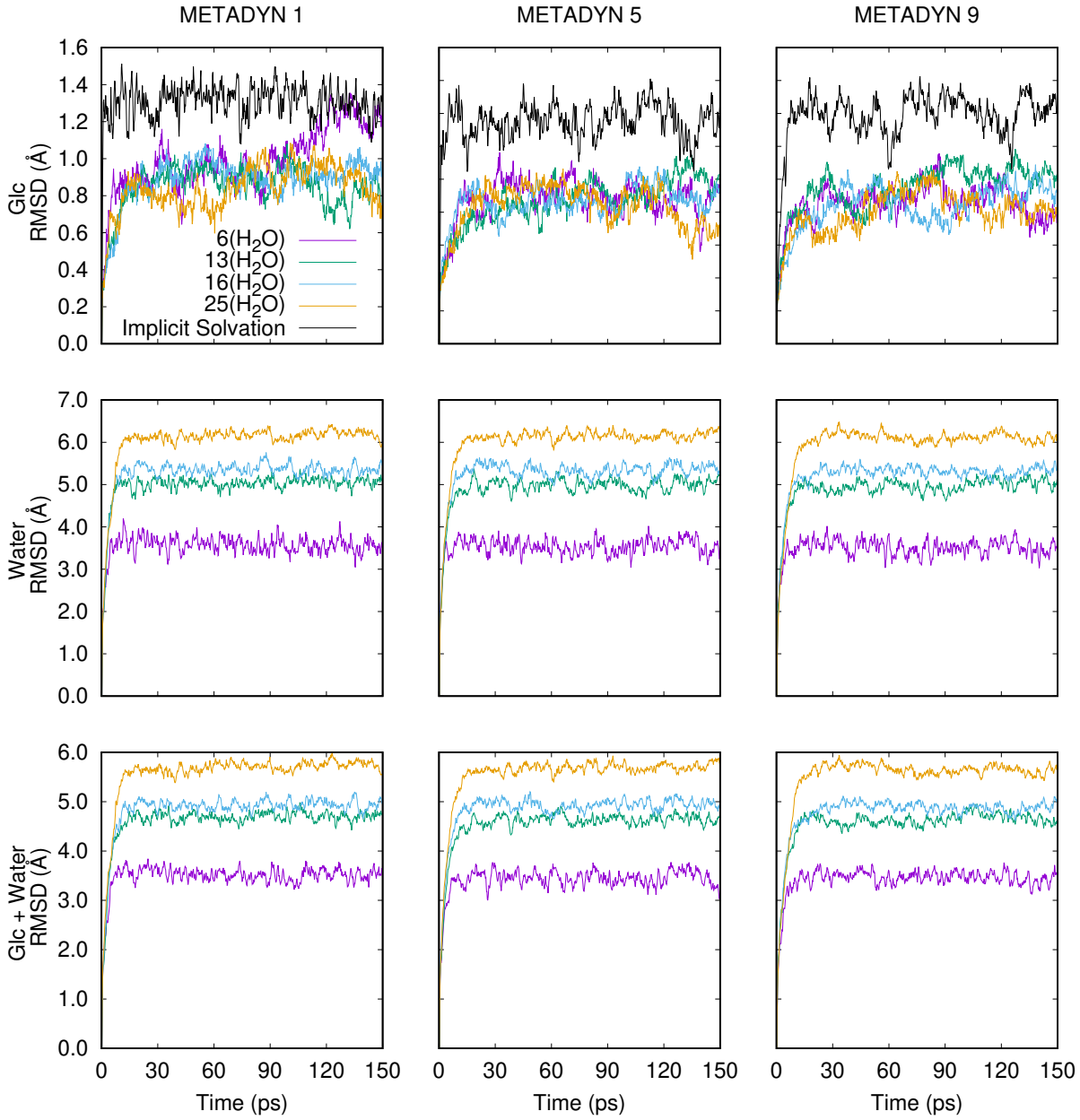


Figure A.13: RMSDs in cartesian space during NCI-MTD simulations for the conformational sampling of glucose. RMSDs are reported for the implicit solvation and microsolvation with 6, 13, 16, and 25 water molecules. The top row shows the RMSDs of solute glucose, the middle row shows the RMSDs of the water molecules for microsolvation, and the bottom row shows the RMSDs for the complete models $Glc(H_2O)_n$. RMSDs were calculated from the average of ten NCI-MTD trajectories in the last iteration for conformational sampling with the CREST program.[110]

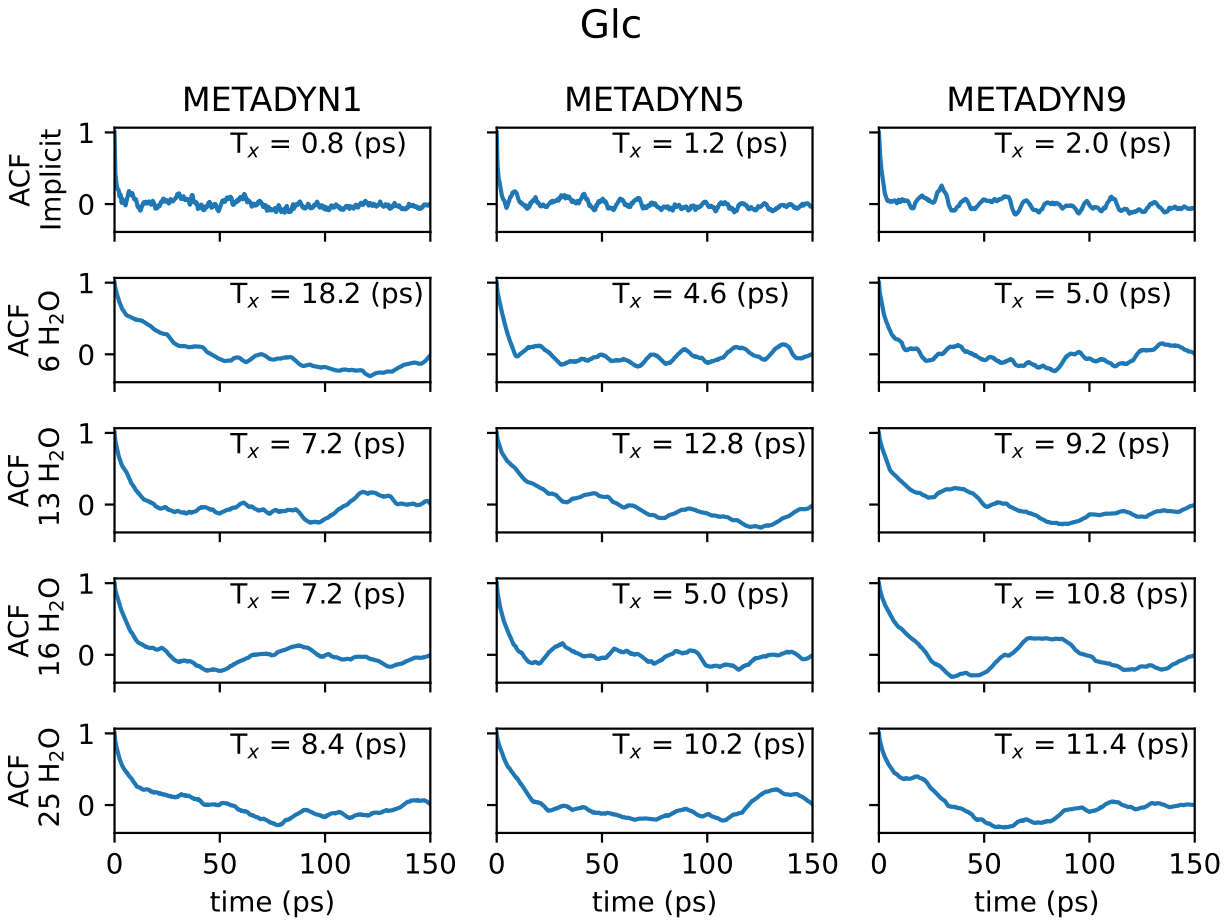


Figure A.14: ACFs for the RMSDs of glucose molecule in the top row of Figure A.10. T_x is the time constant for the corresponding ACF.

Table A.7: Free energies ($G^\circ(X)_{\mathrm{CE}}$, in Hartrees) for the three CEs of Glc and H_2O clusters used for the calculation $\Delta_r G^\circ$ with the setups for setting B, E2, F2, and G2. $(H_2O)_n$ is the number of water molecules for micro-solvation. #conf is the number of conformers for the free energy averages in $G^\circ_{\mathrm{CE}}(X)$. MAD and Range for the three trials reported in kcal/mol.

Setting B									
$(H_2O)_n$	G°_{CE}(Glc)	#conf	MAD	Range	$(H_2O)_n$	$G^\circ_{\mathrm{CE}}(H_2O)$	#conf	MAD	Range
	-686.9815863	98				-305.6881704	616		
0	-686.9815929	96	0.02	0.05	4	-305.6877712	540	0.11	0.27
	-686.9816692	94				-305.687745	855		
	-839.8242346	849				-458.5265375	1317		
2	-839.8241611	723	0.08	0.21	6	-458.526221	1345	0.12	0.35
	-839.8244986	645				-458.5259817	1372		
	-916.244504	551				-534.947155	1183		
3	-916.2434441	76	0.23	0.67	7	-534.9471317	1135	0.15	0.34
	-916.2438789	595				-534.9476701	1151		
	-992.6639847	432				-611.3638494	310		
4	-992.6652002	319	0.37	0.92	8	-611.3637849	803	0.03	0.09
	-992.6637381	258				-611.3639258	306		
	-1069.083532	221				-687.7822127	794		
5	-1069.082669	252	0.24	0.54	9	-687.7825362	773	0.20	0.56
	-1069.083503	171				-687.7831091	728		
	-1145.502712	98				-764.2011008	221		
6	-1145.502682	181	0.47	1.07	10	-764.2002318	14	0.26	0.61
	-1145.50439	152				-764.2012031	591		
	-1221.925882	63				-840.6192831	16		
7	-1221.922992	183	0.72	1.81	11	-840.6190531	30	0.23	0.59
	-1221.925231	98				-840.6199918	552		
	-1451.181987	70				-1069.876062	149		
10	-1451.179441	51	0.63	1.60	14	-1069.87673	215	0.20	0.49
	-1451.181383	62				-1069.87685	233		
	-1680.440526	25				-1299.13975	52		
13	-1680.44109	30	0.37	1.01	17	-1299.135349	42	0.98	2.76
	-1680.43948	50				-1299.137951	21		
	-1909.701595	20				-1528.391961	26		
16	-1909.700772	17	0.96	2.41	20	-1528.394161	25	0.59	1.38
	-1909.697756	29				-1528.393986	24		

Continuation of Table A.7.

Setting E2									
$(H_2O)_n$	G°_{CE}(Glc)	#conf	MAD	Range	$(H_2O)_n$	$G^\circ_{CE}(H_2O)$	#conf	MAD	Range
	-686.938633	101				-305.712455	1052		
0	-686.9386201	105	0.01	0.02	4	-305.7124949	653	0.05	0.12
	-686.9386057	101				-305.7126414	1135		
	-839.7957948	690				-458.5609484	1144		
2	-839.7955482	682	0.21	0.55	6	-458.5610102	1154	0.17	0.40
	-839.7964237	638				-458.561579	1101		
	-916.2218204	428				-534.9859821	1046		
3	-916.2213955	426	0.89	2.13	7	-534.9856874	1067	0.09	0.24
	-916.218427	187				-534.9855969	1043		
	-992.6427896	155				-611.4092451	114		
4	-992.64761	290	1.45	3.49	8	-611.4079388	27	0.35	0.82
	-992.6483589	310				-611.409151	354		
	-1069.072575	131				-687.8340739	546		
5	-1069.072779	138	0.08	0.22	9	-687.8340434	339	0.27	0.61
	-1069.072927	151				-687.833108	20		
	-1145.497392	66				-764.2576812	18		
6	-1145.497058	116	0.90	2.13	10	-764.2575876	24	0.02	0.06
	-1145.500452	118				-764.257602	25		
	-1221.922688	128				-840.6820325	26		
7	-1221.924224	97	0.39	0.96	11	-840.682027	135	0.01	0.02
	-1221.923961	62				-840.6820073	25		
	-1451.197636	51				-1069.955936	220		
10	-1451.197599	27	0.39	0.90	14	-1069.955451	205	0.10	0.30
	-1451.199032	32				-1069.955678	211		
	-1680.47227	19				-1299.232229	59		
13	-1680.473941	17	0.45	1.05	17	-1299.232055	97	0.37	0.90
	-1680.473849	28				-1299.230797	49		
	-1909.750217	9				-1528.506949	70		
16	-1909.745022	10	1.69	4.35	20	-1528.504971	31	0.44	1.24
	-1909.75195	5				-1528.505783	34		

Continuation of Table A.7.

Setting F2									
$(H_2O)_n$	G°_{CE}(Glc)	#conf	MAD	Range	$(H_2O)_n$	$G^\circ_{CE}(H_2O)$	#conf	MAD	Range
	-686.9762544	86				-305.6881704	486		
0	-686.9761749	86	0.02	0.05	4	-305.6877712	893	0.11	0.27
	-686.9762097	86				-305.687745	636		
	-839.8174609	555				-458.5241431	841		
2	-839.8177657	433	0.08	0.19	6	-458.5242071	887	0.02	0.04
	-839.8174698	441				-458.5242013	842		
	-916.2370541	280				-534.9420884	886		
3	-916.2371895	330	0.06	0.18	7	-534.9420572	829	0.01	0.02
	-916.23691	368				-534.942075	876		
	-992.6565011	153				-611.3603088	16		
4	-992.656295	203	0.27	0.67	8	-611.3604666	548	0.06	0.15
	-992.657355	172				-611.3602266	72		
	-1069.075355	53				-687.7789329	7		
5	-1069.074894	162	0.25	0.71	9	-687.7793435	197	0.09	0.26
	-1069.076031	46				-687.7791031	6		
	-1145.494218	98				-764.1973023	24		
6	-1145.494854	79	0.23	0.65	10	-764.1975358	40	0.16	0.43
	-1145.495261	15				-764.1979853	83		
	-1221.912435	50				-840.6149821	92		
7	-1221.912071	65	0.08	0.23	11	-840.6163175	200	0.37	0.84
	-1221.912298	72				-840.6149828	21		
	-1451.170617	16				-1069.870055	143		
10	-1451.17123	23	0.58	1.50	14	-1069.870233	137	0.33	0.80
	-1451.168834	25				-1069.871327	94		
	-1680.427606	9				-1299.130598	28		
13	-1680.425014	12	0.74	1.73	17	-1299.127373	36	0.93	2.14
	-1680.427764	6				-1299.127181	44		
	-1909.681188	6				-1528.383208	11		
16	-1909.681605	4	1.26	2.96	20	-1528.384446	24	0.27	0.78
	-1909.685898	7				-1528.383731	10		

Continuation of Table A.7.

Setting G2									
$(H_2O)_n$	G°_{CE}(Glc)	#conf	MAD	Range	$(H_2O)_n$	$G^\circ_{CE}(H_2O)$	#conf	MAD	Range
	-686.9386042	101				-305.7101591	495		
0	-686.9386208	102	0.01	0.03	4	-305.7102349	906	0.04	0.11
	-686.9386514	100				-305.7100524	645		
	-839.7926732	767				-458.5589312	877		
2	-839.7929273	707	0.07	0.16	6	-458.5587859	926	0.05	0.14
	-839.7928873	654				-458.5590149	887		
	-916.2186532	443				-534.9833805	926		
3	-916.2196328	424	0.24	0.61	7	-534.9837199	867	0.10	0.25
	-916.219375	495				-534.9837804	918		
	-992.6447529	363				-611.4073466	16		
4	-992.6448245	409	0.68	1.55	8	-611.4072098	576	0.11	0.29
	-992.6472213	188				-611.4076744	76		
	-1069.070665	111				-687.8325109	7		
5	-1069.071289	137	1.69	4.00	9	-687.8318995	214	0.24	0.70
	-1069.06492	139				-687.8314003	6		
	-1145.495147	184				-764.2556803	24		
6	-1145.49559	111	0.32	0.86	10	-764.2558679	42	0.06	0.14
	-1145.49652	110				-764.2559027	90		
	-1221.922332	62				-840.6795914	99		
7	-1221.925187	52	0.89	2.22	11	-840.6805404	220	0.37	1.05
	-1221.921656	37				-840.6812684	21		
	-1451.197392	51				-1069.9543	158		
10	-1451.196563	27	0.50	1.39	14	-1069.954769	154	0.17	0.49
	-1451.198778	20				-1069.953993	109		
	-1680.472887	36				-1299.231562	59		
13	-1680.471818	20	0.97	2.52	17	-1299.228757	48	0.74	1.76
	-1680.475829	3				-1299.231264	63		
	-1909.748882	6				-1528.505475	14		
16	-1909.748225	7	0.53	1.40	20	-1528.504323	36	0.31	0.72
	-1909.746658	25				-1528.505417	20		

A.7 Distribution of Reaction-Free Energies Relative to the Number of Trials $G^{\circ}_{\mathrm{CE},i}(X)$

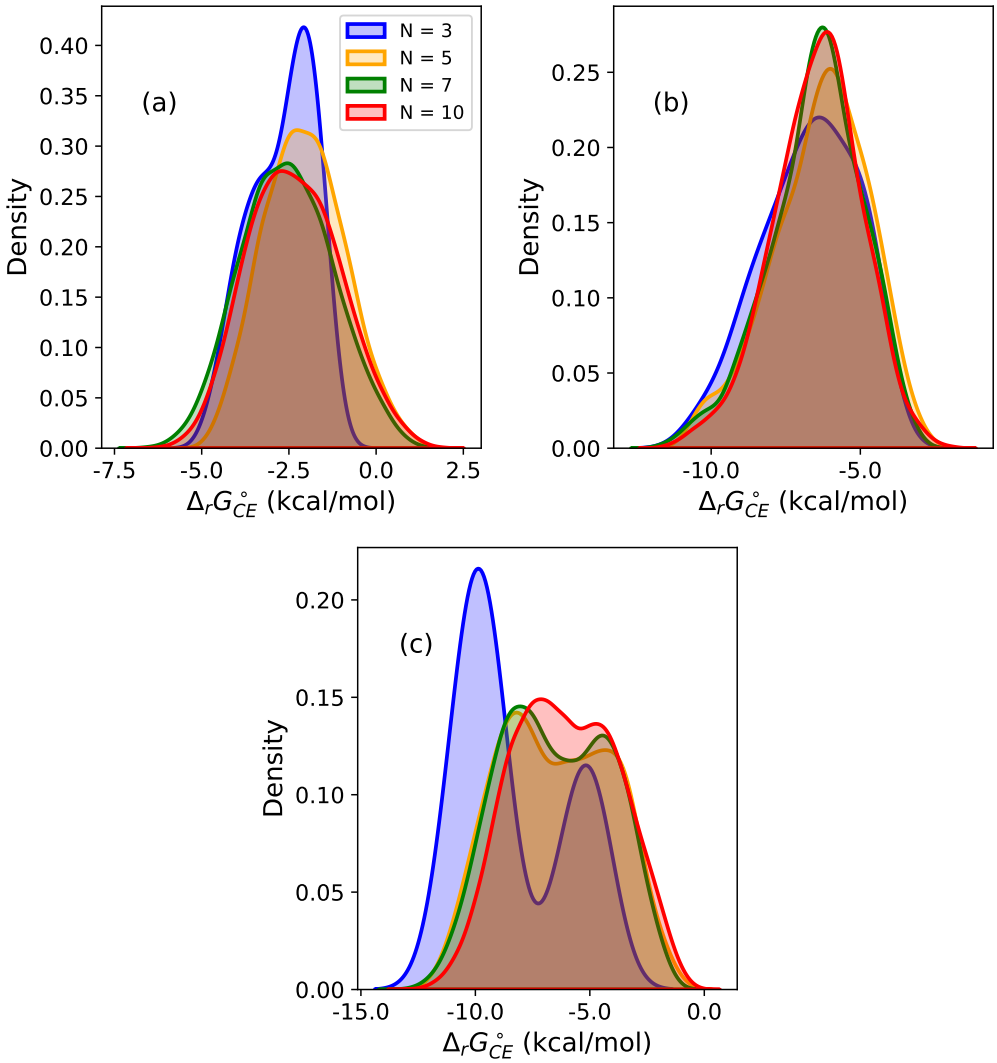


Figure A.15: Kernel density estimation for the reaction-free energies distributions $\{\Delta_r G^{\circ}_{\mathrm{CE},x}\}$ of reactions (a) **DBA-1**$^{-2}$ + Glc $\overset{K}{\rightleftharpoons}$ **DBA-1.Glc**$^{-2}$ + 4 H_2O, (b) **DBA-2** + Glc $\overset{K}{\rightleftharpoons}$ **DBA-2.Glc** + 4 H_2O, and (c) **DBA-1**$^{-2}$ + Gal $\overset{K}{\rightleftharpoons}$ **DBA-1.Gal**$^{-2}$ + 4 H_2O. N is the number of trials for the molecular free energies $G^{\circ}_{\mathrm{CE},i}(X)$, for reactants and products, used in eq 3.2.

Appendix B

Molecular Fragments Used in the GB-GA Optimization Process

Table B.1: SMILES representations for aBAs used in mutation operations of the graph-based genetic algorithm. The boronic acid group and the linker fragment are located at the 1,2-positions on the phenyl ring. [98*] represents the dummy atom for binding the boronic acid group, and [96*] represents the dummy atom for binding the linker on the glucose side of sucrose. For the fructose side, the same aBAs are used, with [99*] indicating the boronic acid binding site and [97*] indicating the linker binding site.

SMILES	pK_a	Parent PBA	Reference
C-O-C1=CC=C(-[96*])C(-[98*])=C1	9.0	2-Methoxyphenylboronic acid	49
N-C1=CC=C(-[96*])C(-[98*])=C1	8.9	3-Aminophenylboronic acid	49
N-C1=CC=CC(-[98*])=C1-[96*]	8.9	3-Aminophenylboronic acid	49
[96*]-C1=CC=CC=C1-[98*]	8.8	Phenylboronic acid	49
FC1=CC=C(-[98*])C(-[96*])=C1	8.6	4-Fluorophenylboronic acid	49
ClC1=CC(Cl)=C(-[98*])C(-[96*])=C1	8.5	2,4-Diclorophenylboronic acid	49
BrC1=CC=C(-[98*])C(-[96*])=C1	8.8	4-Bromophenylboronic acid	49
N-C-C1=CC=C(-[98*])C(-[96*])=C1	8.3	4-Aminomethylphenylboronic acid	49
[96*]-C1=CC=NC=C1-[98*]	8.1	3-Pyridinylboronic acid	49
[96*]-C1=CN=CC=C1-[98*]	8.0	4-Pyridinylboronic acid	49

Continuation of Table B.1

SMILES	pK_a	Parent PBA	Reference
O-C(=O)-C1=CC=C(-[98*])C(-[96*])=C1	8.0	4-Carboxyphenylboronic acid	49
C-C(=O)-C1=CC=CC(-[98*])=C1-[96*]	8.0	3-Acetophenylboronic acid	49
C-C(=O)-C1=CC=C(-[96*])C(-[98*])=C1	8.0	3-Acetophenylboronic acid	49
FC1=CC(-[96*])=C(-[98*])C=C1Cl	7.8	3-Chloro-4-fluorophenylboronic acid	49
FC1=CC=C(-[98*])C(-[96*])=C1Cl	7.8	3-Chloro-4-fluorophenylboronic acid	49
C-C(=O)-C1=CC=C(-[98*])C(-[96*])=C1	7.7	4-Acetylphenylboronic acid	49
[96*]-C1=CC(-C=O)=CC=C1-[98*]	8.0	3-Acetophenylboronic acid	49
FC1=CC(-[96*])=C(-[98*])C=C1Cl	7.8	3-Chloro-4-fluorophenylboronic acid	49
FC1=CC=C(-[98*])C(-[96*])=C1Cl	7.8	3-Chloro-4-fluorophenylboronic acid	49
[96*]-C1=C(-[98*])C=CC=C1-C=O	7.8	3-Formylphenylboronic acid	49
[96*]-C1=CC=C(-C=O)C=C1-[98*]	7.8	3-Formylphenylboronic acid	49
C-C(=O)-C1=CC=C(-[98*])C(-[96*])=C1	7.7	4-Acetylphenylboronic acid	49
C(=O)-C1=CC=C(-[98*])C(-[96*])=C1	7.6	4-Formylphenylboronic acid	49
FC1=CC(F)=C(-[98*])C(-[96*])=C1	7.6	2,4-Difluorophenylboronic acid	49
[O-]-[N+](=O)-C1=CC=C(-[96*])C(-[98*])=C1	7.1	3-Nitrophenylboronic acid	49
[O-]-[N+](=O)-C1=CC=CC(-[98*])=C1-[96*]	7.1	3-Nitrophenylboronic acid	49
FC1=CC=C(F)C(-[98*])=C1-[96*]	7.0	2,5-Diflourophenylboronic acid	49
FC1=CC(F)=C(-[98*])C(-[96*])=C1F	6.8	3,4,5-Trifluorophenylboronic acid	49
FC1=CC(-[96*])=C(-[98*])C(F)=C1F	6.8	2,3,4-Trifluorophenylboronic acid	49
FC1=CC=C(-C=O)C(-[98*])=C1-[96*]	6.72	5-fluoro-2-formylphenylboronic acid	210
FC(F)(F)C1=CC=C(-C=O)C(-[98*])=C1-[96*]	6.72	5-Trifluoromethyl-2-formylphenylboronic acid	210
[O-]-[N+](=O)-C1=CC=C(F)C(-[98*])=C1-[96*]	6.0	2-Fluoro-5-nitrophenylboronic acid	49
[96*]-C1=CC=CC(-C=O)=C1-[98*]	7.31	2-Formylphenilboronic acid	210
FC1=CC=C(-[96*])C(-[98*])=C1	7.5	3-Fluorophenilboronic acid	210
FC1=CC=CC(-[98*])=C1-[96*]	7.5	3-Fluorophenilboronic acid	210
FC(F)(F)C1=CC=C(-[96*])C(-[98*])=C1	7.85	3-Metiltrifluorophenilboronic acid	210
[96*]-C1=CC(=CC=C1-[98*])S(=O)(=O)C-C-C=C	7.1	4-(3-butenylsulfonyl)phenylboronic acid (BSPBA)	211
[96*]-C1=CC(=CC=C1-[98*])S(=O)(=O)N-C-C=C	7.4	4-(N-allylsulfamoyl)phenylboronic acid (BSPBA)	211
N-C1=CC(-[98*])=C(-[96*])C=C1-[N+](-[O-])=O	7.1	(3-amino-4nitrophenyl)boronic acid	212
N-C1=C(-[96*])C(-[98*])=CC=C1-[N+](-[O-])=O	7.1	(3-amino-4nitrophenyl)boronic acid	212
[96*]-C1=C(-[98*])C=CC=C1-N-C(=O)-C=C	8.2	3-acrylamidophenylboronic acid (AAPBA)	213
[96*]-C1=CC=C(-N-C(=O)-C=C)C=C1-[98*]	8.2	3-acrylamidophenylboronic acid (AAPBA)	213
[96*]-C1=CC(=CC=C1-[98*])-C(=O)-N-C	7.8	4-(1,6-dioxo-2,5-diaza-7-oxamyl-)-phenylboronic acid (DDOPPA)	213
FC1=CC(-[96*])=C(-[98*])C(F)=C1-C=O	6.5	2,4-difluoro-3-formyl-phenylboronicacid (DFFPBA)	213
FC1=NC=CC(-[96*])=C1-[98*]	7.1	2-fluoro-3-pyridylboronic acid or 2F-3-PyBA	214
FC1=CC(-[96*])=C(-[98*])C=N1	7.0	2-fluoro-5-pyridylboronic acid or 2F-5-PyBA	214
FC1=CC=C(-[98*])C(-[96*])=N1	7.0	2-fluoro-5-pyridylboronic acid or 2F-5-PyBA	214

Continuation of Table B.1

SMILES	pK_a	Parent PBA	Reference
[96*]-C1=NC=NC=C1-[98*]	6.2	pyrimidine-5-boronic acid	215
[96*]-C1=CC=NC=C1-[98*]	4.4	3-pyridylboronic acid	215
[96*]-C1=NC=CC=C1-[98*]	4.4	3-pyridylboronic acid	215
O-C(=O)-C1=CC(-[96*])=C(-[98*])C=N1	4.2	5-boronopicolinic acid	215
O-C(=O)-C1=CC=C(-[98*])C(-[96*])=N1	4.2	5-boronopicolinic acid	215
C-N-C(=O)-C1=CC(-[96*])=C(-[98*])C=N1	4.2	(6-propylcarbamoyl)pyridine-3-boronic acid	215
C-N-C(=O)-C1=CC=C(-[98*])C(-[96*])=N1	4.2	(6-propylcarbamoyl)pyridine-3-boronic acid	215
O-C(=O)-C-[N+]1=CC=CC(-[98*])=C1-[96*]	4.4	3-borono-1-(carboxymethyl)pyridine	215
O-C(=O)-C-[N+]1=CC=C(-[96*])C(-[98*])=C1	4.4	3-borono-1-(carboxymethyl)pyridine	215
C-C(=O)-N-C1=CC=C(-[96*])C(-[98*])=C1	8.3	3-propionamidophenylboronic acid	215
C-C(=O)-N-C1=CC=CC(-[98*])=C1-[96*]	8.3	3-propionamidophenylboronic acid	215
CS(=O)(=O)C1=CC=C(-[98*])C=C1-[96*]	7.1	4-(methylsulfonyl)benzeneboronic acid	215
FC1=NC=CC(-[96*])=C1-[98*]	6.3	2-Fluoro-3-pyridyl boronic acid	215

Table B.2: SMILES representations for aBAs used in mutation operations of the graph-based genetic algorithm. The boronic acid group and the linker fragment are located at the 1,3-positions on the phenyl ring. [98*] represents the dummy atom for binding the boronic acid group, and [96*] represents the dummy atom for binding the linker on the glucose side of sucrose. For the fructose side, the same aBAs are used, with [99*] indicating the boronic acid binding site and [97*] indicating the linker binding site.

SMILES	pK_a	Parent PBA	Reference
C-O-C1=CC=C(-[96*])C=C1-[98*]	9.0	2-Methoxyphenylboronic acid	49
N-C1=CC(-[96*])=CC(-[98*])=C1	8.9	3-Aminophenylboronic acid	49
C1=CC(-[96*])=CC(-[98*])=C1	8.8	Phenylboronic acid	49
FC1=CC=C(-[98*])C=C1-[96*]	8.6	4-Fluorophenylboronic acid	49
ClC1=CC(Cl)=C(-[98*])C=C1-[96*]	8.5	2,4-Diclorophenylboronic acid 2	49
ClC1=CC=C(-[98*])C(Cl)=C1-[96*]	8.5	2,4-Diclorophenylboronic acid 2	49
Br-C1=CC=C(-[98*])C=C1-[96*]	8.8	4-Bromophenylboronic acid	49
N-C-C1=CC=C(-[98*])C=C1-[96*]	8.3	4-Aminomethylphenylboronic acid	49
[96*]-C1=CN=CC(-[98*])=C1	8.1	3-Pyridinylboronic acid	49
[96*]-C1=CC(-[98*])=CC=N1	8.0	4-Pyridinylboronic acid	49
O-C(=O)-C1=CC=C(-[98*])C=C1-[96*]	8.0	4-Carboxyphenylboronic acid	49
C-C(=O)-C1=CC(-[96*])=CC(-[98*])=C1	8.0	3-Acetophenylboronic acid	49
FC1=C(Cl)C=C(-[98*])C=C1-[96*]	7.8	3-Chloro-4-fluorophenylboronic acid	49

Continuation of Table B.2

SMILES	pK_a	Parent PBA	Reference
[96*]-C1=CC(-[98*])=CC(-C=O)=C1	7.8	3-Formylphenylboronic acid	49
C-C(=O)-C1=CC=C(-[98*])C=C1-[96*]	7.7	4-Acetylphenylboronic acid	49
[96*]-C1=CC(-[98*])=CC=C1-C=O	7.6	4-Formylphenylboronic acid	49
FC1=CC(F)=C(-[98*])C=C1-[96*]	7.6	2,4-Difluorophenylboronic acid 1	49
FC1=CC=C(-[98*])C(F)=C1-[96*]	7.6	2,4-Difluorophenylboronic acid 2	49
[O-]-[N+](=O)-C1=CC(-[96*])=CC(-[98*])=C1	7.1	3-Nitrophenylboronic acid	49
FC1=CC(-[98*])=C(F)C(-[96*])=C1F	6.8	3,4,5-Trifluorophenylboronic acid	49
FC1=C(F)C(-[96*])=CC(-[98*])=C1F	6.8	2,3,4-Trifluorophenylboronic acid	49
FC1=CC(-[98*])=C(F)C(-[96*])=C1F	6.7	2,4,5-Trifluorophenylboronic acid	49
[98*]-C1C(-C[NH+](C)C)=CC(-[96*])=CC1	5.3	2-Dimethylaminomethylphenylboronic acid (DAPBA)	49
[O-]-[N+](=O)-C1=CC(-[96*])=C(F)C(-[98*])=C1	6.0	2-Fluoro-5-nitrophenylboronic acid	49
C-C1=CC(-[98*])=C[N+](-[#6]-[96*])=C1	4.4	5-Methylpyridine-3-boronic acid	49
[96*]-C-[N+]1=CC=CC(-[98*])=C1	4.4	3-Methylpyridineboronic acid	49
[96*]-C1=CC=C(-C=O)C(-[98*])=C1	7.31	2-Formylphenilboronic acid	210
FC1=CC(-[96*])=CC(-[98*])=C1	7.5	3-Fluorophenilboronic acid	210
FC(F)(F)C1=CC(-[96*])=CC(-[98*])=C1	7.85	3-Metiltrifluorophenilboronic acid	210
N-C1=CC(-[98*])=CC(-[96*])=C1-[N+](-[O-])=O	7.1	(3-amino-4nitrophenyl)boronic acid	210
[96*]-C1=CC(-[98*])=CC(-N-C(=O)-C=C)=C1	8.2	3-acrylamidophenylboronic acid (AAPBA)	213
[96*]-C1=CC(-[98*])=CC=C1-C(=O)-N-C	7.8	4-(1,6-dioxo-2,5-diaza-7-oxamyl-)-phenylboronic acid (DDOPBA)	216
FC1=C(-[96*])C=C(-[98*])C(F)=C1-C=O	6.5	2,4-difluoro-3-formyl-phenylboronicacid (DFFPBA)	213
FC1=NC=C(-[96*])C=C1-[98*]	7.1	2-fluoro-3-pyridylboronic acid or 2F-3-PyBA	214
FC1=NC=C(-[98*])C=C1-[96*]	7.0	2-fluoro-5-pyridylboronic acid or 2F-5-PyBA	214
[96*]-C-[N+]1=CN=CC(-[98*])=C1	6.2	pyrimidine-5-boronic acid	215
[96*]-C1=CN=CC(-[98*])=C1	4.4	3-pyridylboronic acid	215
O-C(=O)-C1=NC=C(-[98*])C=C1-[96*]	4.2	5-boronopicolinic acid	215
O-C(=O)-C1=CC=C(-[98*])C=[N+]1-C-[96*]	4.2	5-boronopicolinic acid	215
C-N-C(=O)-C1=NC=C(-[98*])C=C1-[96*]	4.2	(6-propylcarbamoyl)pyridine-3-boronic acid	215
C-N-C(=O)-C1=CC=C(-[98*])C=[N+]1-C-[96*]	4.2	(6-propylcarbamoyl)pyridine-3-boronic acid	215
O-C(=O)-C-[N+]1=CC(-[96*])=CC(-[98*])=C1	4.4	3-borono-1-(carboxymethyl)pyridine	215
C-C(=O)-N-C1=CC(-[96*])=CC(-[98*])=C1	8.3	3-propionamidophenylboronic acid	215
CS(=O)(=O)C1=CC=C(-[98*])C=C1-[96*]	7.1	4-(methylsulfonyl)benzeneboronic acid	215
FC1=NC=C(-[96*])C=C1-[98*]	6.3	2-Fluoro-3-pyridyl boronic acid	215
FC1=C(-[98*])C=CC=[N+]1-C-[96*]	6.3	2-Fluoro-3-pyridyl boronic acid	215

Table B.3: SMILES representations for aBAs used in mutation operations of the graph-based genetic algorithm. The boronic acid group and the linker fragment are located at the 1,4-positions on the phenyl ring. [98*] represents the dummy atom for binding the boronic acid group, and [96*] represents the dummy atom for binding the linker on the glucose side of sucrose. For the fructose side, the same aBAs are used, with [99*] indicating the boronic acid binding site and [97*] indicating the linker binding site.

SMILES	pK_a	Parent PBA	Reference
N-C1=CC(-[98*])=CC=C1-[96*]	8.9	3-Aminophenylboronic acid	49
[96*]-C1=CC=C(-[98*])C=C1	8.8	Phenylboronic acid	49
[96*]-C1=CC=C(-[98*])C=N1	8.1	3-Pyridinylboronic acid	49
C-C(=O)-C1=CC(-[98*])=CC=C1-[96*]	8.0	3-Acetophenylboronic acid	49
[96*]-C1=CC=C(-[98*])C=C1-C=O	7.8	3-Formylphenylboronic acid	49
[O-]-[N+](=O)-C1=CC(-[98*])=CC=C1-[96*]	7.1	3-Nitrophenylboronic acid 1	49
FC1=CC(-[98*])=C(F)C=C1-[96*]	7.0	2,5-Diflourophenylboronic acid	49
[O-]-[N+](=O)-C1=CC(-[98*])=C(F)C=C1-[96*]	6.0	2-Fluoro-5-nitrophenylboronic acid	49
[96*]-C[N+]1=CC=C(-[98*])C=C1	4.4	4-Methylpyridineboronic acid	49
[96*]-C1=CC=C(-[98*])C(-C=O)=C1	7.31	2-Formylphenilboronic acid	210
FC1=CC(-[98*])=CC=C1-[96*]	7.5	3-Fluorophenilboronic acid	210
FC(F)(F)C1=CC(-[98*])=CC=C1-[96*]	7.85	3-Metiltrifluorophenilboronic acid	210
[96*]-C1=CC=C(-[98*])C=C1-N-C(=O)-C	8.2	3-acrylamidophenylboronic acid (AAPBA)	213
[96*]-C1=CC=C(-[98*])C=C1-C(=O)-N-C	7.8	4-(1,6-dioxo-2,5-diaza-7-oxamyl-)-phenylboronic acid (DDOPPA)	216
FC1=NC(-[96*])=CC=C1-[98*]	7.1	2-fluoro-3-pyridylboronic acid or 2F-3-PyBA	214
[96*]-C1=NC=C(-[98*])C=N1	6.2	pyrimidine-5-boronic acid	215
[96*]-C1=CC=C(-[98*])C=N1	4.4	3-pyridylboronic acid	215
O-C(=O)-C-[N+]1=CC(-[98*])=CC=C1-[96*]	4.4	3-borono-1-(carboxymethyl)pyridine	215
C-C-C(=O)-N-C1=CC(-[98*])=CC=C1-[96*]	8.3	3-propionamidophenylboronic acid	215
FC1=NC(-[96*])=CC=C1-[98*]	6.3	2-Fluoro-3-pyridyl boronic acid	215

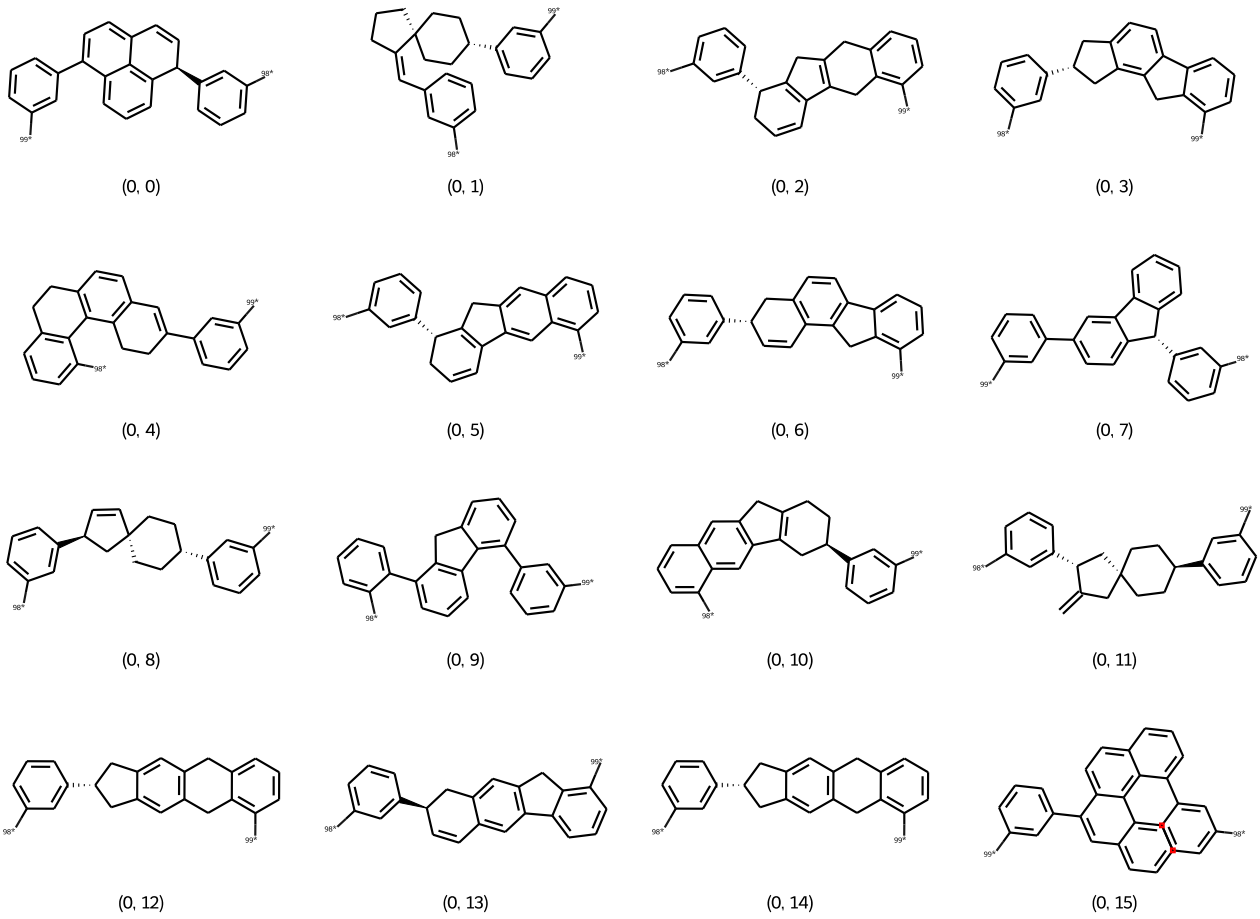
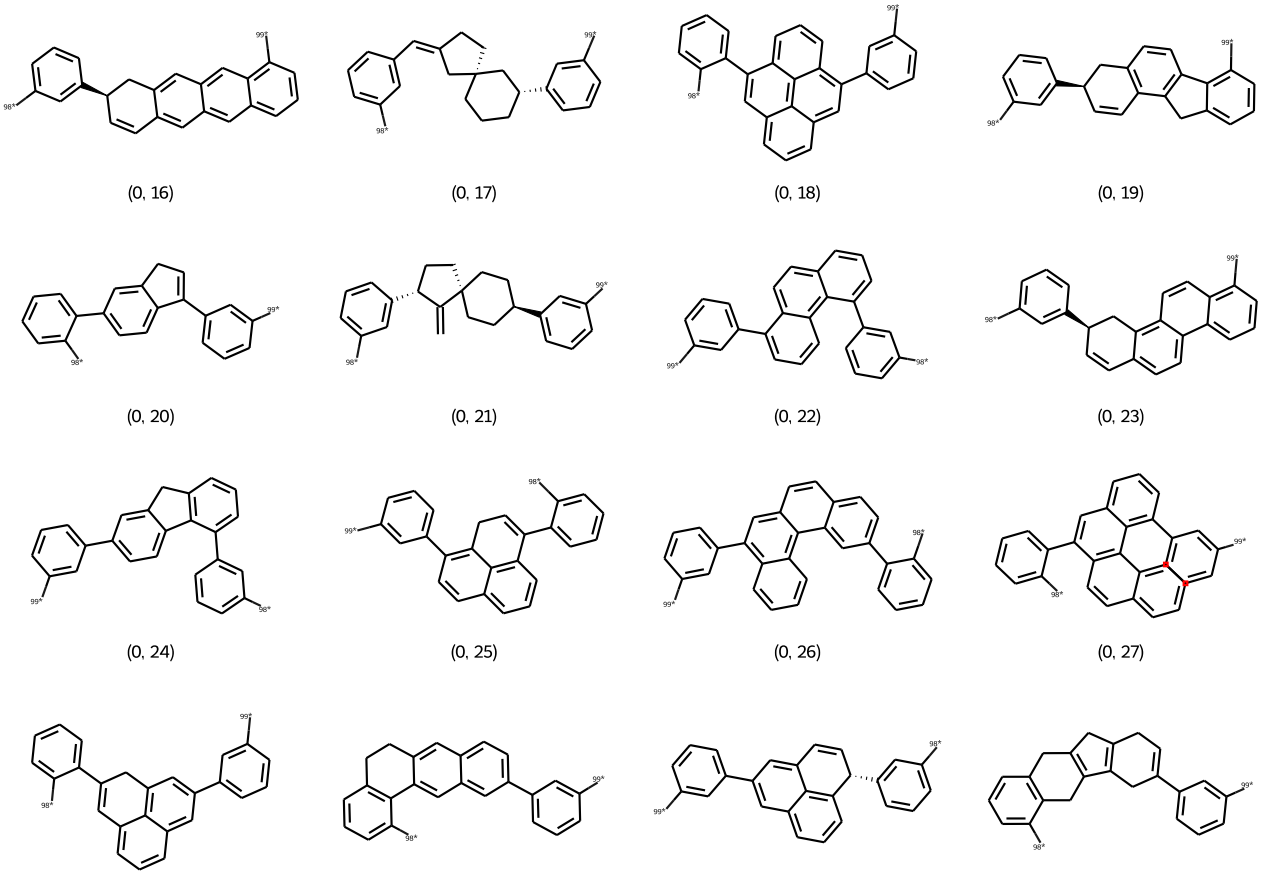


Figure B.1: Initial population of DBA tweezers used for parameter screening of mutation rate (MUT) and selection pressure (η) reported in Figure 4.9. The first number in parentheses represents the tweezer generation, while the second number denotes the nominal index within that generation.



Continuation of Figure B.1.

Table B.4: Sample outputs of the GB-GA optimization from a single run with a mutation rate of 0.5 and η parameter set to 1.5. In the idx column, the first number in parentheses represents the tweezer generation, while the second number denotes the nominal index within that generation.

Generation 0						
idx	Normalized Fitness	Score	Free Energy	LogP	pKa max	SA score
(0, 27)	0.046875	0.654569	-13.0495	6	8.8	4.41327
(0, 17)	0.0458669	0.685037	-9.04894	4.44946	8.8	4.43722
(0, 8)	0.0448589	0.689096	-8.26355	4.10542	8.8	4.37945
(0, 28)	0.0438508	0.699889	-9.35914	5.30356	8.8	4.48978
(0, 3)	0.0428427	0.713063	-7.35507	4.26877	8.8	4.7205
(0, 9)	0.0418347	0.718293	-8.95396	6	8.8	4.27331
(0, 2)	0.0408266	0.721613	-7.01572	4.74779	8.8	4.26638
(0, 21)	0.0398185	0.722788	-6.26016	4.24434	8.8	4.27697
(0, 11)	0.0388105	0.723111	-6.86475	4.69472	8.8	4.28095
(0, 13)	0.0378024	0.723484	-7.64159	4.77044	8.8	4.95416
(0, 24)	0.0367944	0.724277	-7.54187	5.50934	8.8	3.92002
(0, 20)	0.0357863	0.732132	-7.07966	5.41759	8.8	4.06635
(0, 22)	0.0347782	0.73332	-7.34363	5.63392	8.8	4.10466
(0, 15)	0.0337702	0.735245	-7.83123	6	8.8	4.20394
(0, 0)	0.0327621	0.737092	-8.06247	5.69721	8.8	4.94223
(0, 10)	0.031754	0.737346	-6.6616	4.74454	8.8	4.87199
(0, 30)	0.030746	0.738212	-7.42881	5.27553	8.8	4.95845
(0, 12)	0.0297379	0.741645	-6.61041	4.9748	8.8	4.77338
(0, 14)	0.0287298	0.742358	-6.56619	4.9748	8.8	4.77338
(0, 23)	0.0277218	0.74383	-6.19544	5.59692	8.8	3.67342
(0, 6)	0.0267137	0.745319	-5.79485	4.36918	8.8	5.01211
(0, 31)	0.0257056	0.748743	-5.79647	4.54131	8.8	4.98959
(0, 5)	0.0246976	0.750366	-4.31387	4.31393	8.8	3.95538
(0, 18)	0.0236895	0.75132	-7.0688	6	8.8	4.43055
(0, 1)	0.0226815	0.755161	-4.54177	4.85045	8.8	3.74828
(0, 7)	0.0216734	0.761024	-5.59789	5.46929	8.8	4.29699
(0, 29)	0.0206653	0.763201	-5.8217	5.55012	8.8	4.53639
(0, 16)	0.0196573	0.772704	-5.06991	5.18965	8.8	4.85965
(0, 19)	0.0186492	0.793215	-2.86063	5.18311	8.8	3.96105
(0, 4)	0.0176411	0.867333	1	5.13683	8.8	4.73372
(0, 25)	0.0166331	0.869303	1	5.41812	8.8	4.47688
(0, 26)	0.015625	0.87734	1	6	8.8	4.18325

Continuation of Table B.4.

Generation 25						
idx	Normalized Fitness	Score	Free Energy	LogP	pKa max	SA score
(20, 3)	0.046875	0.467479	-12.7759	-2.17232	7.1	6.00533
(24, 15)	0.0458669	0.490081	-9.73297	-3	7.1	5.52016
(25, 2)	0.0448589	0.497161	-9.75619	-3	7.1	5.96743
(24, 6)	0.0438508	0.502278	-12.7497	-2.66582	9	6.28297
(19, 0)	0.0428427	0.502451	-12.8508	-1.23328	7.85	5.95977
(22, 15)	0.0418347	0.505971	-8.69816	-1.8401	7.1	3.92559
(23, 5)	0.0408266	0.506312	-13.5752	-0.567065	7.85	6.00413
(16, 9)	0.0398185	0.511052	-11.8055	-1.71707	7.85	6.10926
(17, 8)	0.0388105	0.513264	-8.81789	-3	7.1	6.02557
(9, 11)	0.0378024	0.513288	-18.369	3.52332	8.8	4.38664
(14, 11)	0.0367944	0.514827	-8.22543	-3	8.1	4.26031
(22, 2)	0.0357863	0.517426	-18.5468	3.63551	8.9	4.52877
(18, 13)	0.0347782	0.520257	-7.7039	-2.9592	8.1	4.02699
(25, 13)	0.0337702	0.522515	-12.1893	-0.120845	7.1	6.0044
(24, 14)	0.0327621	0.523113	-9.54529	-1.82026	7.1	5.74748
(15, 0)	0.031754	0.525388	-7.89536	-2.89658	7.1	5.72238
(23, 2)	0.030746	0.527234	-7.62529	-3	7.1	5.70969
(23, 0)	0.0297379	0.532175	-11.1067	-0.157965	7.1	5.58578
(12, 5)	0.0287298	0.533921	-11.5595	-0.0587754	7.1	5.99653
(18, 10)	0.0277218	0.536359	-12.8345	0.920216	8.1	4.78573
(19, 2)	0.0267137	0.540548	-12.3331	0.754105	7.1	6.05894
(19, 12)	0.0257056	0.548067	-14.1943	2.93991	8.2	3.98251
(16, 14)	0.0246976	0.548422	-10.414	-0.649274	7.85	5.58113
(23, 3)	0.0236895	0.55917	-7.89695	-1.61226	8.1	4.75269
(12, 3)	0.0226815	0.560998	-10.5149	-0.113545	9	4.2404
(18, 14)	0.0216734	0.56193	-3.97001	-2.9053	7	4.25635
(15, 8)	0.0206653	0.567294	-8.37551	-1.47258	7.85	5.83845
(12, 0)	0.0196573	0.567889	-10.8129	1.49465	7.85	4.2766
(24, 7)	0.0186492	0.567991	-8.76902	-1.41869	7.85	6.1892
(16, 11)	0.0176411	0.568435	-6.14108	-3	8	5.58828
(19, 14)	0.0166331	0.570009	-12.6703	0.417151	9	6.15936
(16, 5)	0.015625	0.572252	-6.91309	-1.43063	7.1	5.62901

Continuation of Table B.4.

Generation 50						
idx	Normalized Fitness	Score	Free Energy	LogP	pKa max	SA score
(48, 6)	0.046875	0.31533	-23.203	-0.51022	7.1	4.75088
(43, 1)	0.0458669	0.396912	-15.4446	-2.2352	7.1	4.43769
(37, 2)	0.0448589	0.398732	-16.4382	-1.29389	7.1	4.25329
(31, 13)	0.0438508	0.414838	-14.8665	0.671077	4.4	4.55024
(50, 0)	0.0428427	0.42919	-12.7296	-2.79673	7.1	4.49559
(42, 4)	0.0418347	0.432232	-15.1218	-1.11301	7.1	4.74826
(26, 12)	0.0408266	0.437965	-13.8274	-2.40936	7.1	5.56802
(36, 13)	0.0398185	0.441334	-13.6255	-1.33277	7.1	4.13934
(32, 10)	0.0388105	0.444872	-13.0904	-3	7.1	6.05676
(43, 9)	0.0378024	0.449547	-13.8215	1.30615	4.4	4.77465
(31, 7)	0.0367944	0.452782	-10.9707	-2.82229	7.1	4.24306
(42, 10)	0.0357863	0.458655	-12.3077	-3	7.1	6.12631
(49, 3)	0.0347782	0.458733	-11.4996	-3	7.1	5.34898
(37, 0)	0.0337702	0.460483	-12.8996	-1.15248	7.1	4.34544
(48, 1)	0.0327621	0.460659	-12.2846	-3	8.6	4.2956
(27, 12)	0.031754	0.463091	-12.9325	-2.29572	7.85	5.09379
(20, 3)	0.030746	0.467479	-12.7759	-2.17232	7.1	6.00533
(28, 12)	0.0297379	0.468477	-11.2619	-3	7.1	5.70352
(39, 1)	0.0287298	0.469307	-11.0622	-2.6625	7.1	5.11008
(30, 15)	0.0277218	0.473703	-10.5133	-3	7.1	5.2926
(39, 7)	0.0267137	0.475817	-10.503	-2.96783	7.1	5.36664
(42, 5)	0.0257056	0.47623	-13.8765	-1.01397	7.1	6.05092
(50, 11)	0.0246976	0.479101	-12.073	-0.893599	7.1	4.31743
(43, 13)	0.0236895	0.479951	-10.8303	-2.04225	7.1	4.69732
(49, 1)	0.0226815	0.481356	-12.0331	-1.4865	7.1	5.20463
(42, 0)	0.0216734	0.486033	-9.94368	-2.03075	7.1	4.18889
(45, 13)	0.0206653	0.48652	-11.1632	-1.22316	7.1	4.32147
(50, 7)	0.0196573	0.488568	-16.9758	3.08092	7.1	4.33068
(30, 6)	0.0186492	0.489339	-8.72475	-2.71949	7.1	4.12596
(24, 15)	0.0176411	0.490081	-9.73297	-3	7.1	5.52016
(49, 0)	0.0166331	0.4903	-8.45772	-3	7.1	4.2992
(42, 11)	0.015625	0.491424	-11.9894	-0.321814	7.1	4.21346

Appendix C

Reaction Mechanism of PBA with Sucrose under Reaction Condition pH $<$ p$K_{a,\mathrm{PBA}}$

Table C.1: Reaction rate coefficients ($k(T)$) and half-life times ($t_{1/2}$) for the reactions of PBA with sucrose under conditions of pH $<$ p$K_{a,\mathrm{PBA}}$. $\Delta_r G^{\circ,\ddagger}$ is 27.2 kcal/mol[a] for the glucose-side reaction and 28.6 kcal/mol[a] for the fructose-side reaction.

	Glucose-side		Fructose-side	
$T(^\circ C)$	$k(T)(s^{-1}M^{-1})$	$t_{1/2}\,(min)$	$k(T)(s^{-1}M^{-1})$	$t_{1/2}\,(min)$
25	7.5×10^{-8}	1.5×10^{8}	7.0×10^{-9}	1.6×10^{9}
30	1.6×10^{-7}	7.3×10^{7}	1.6×10^{-8}	7.4×10^{8}
35	3.3×10^{-7}	3.5×10^{7}	3.4×10^{-8}	3.4×10^{8}
40	6.7×10^{-7}	1.7×10^{7}	7.1×10^{-8}	1.6×10^{8}
50	2.6×10^{-6}	4.4×10^{6}	2.9×10^{-7}	3.9×10^{7}

[a] $k(T)$ calculated with the Eyring equation.[178] [b] $t_{1/2} = \ln(2)/k(T)[\mathrm{PBA}]_0$, where $[\mathrm{PBA}]_0 = 1.0 \times 10^{-3}\,\mathrm{M} << [\mathrm{Suc}]_0$.[180] [c] $\Delta_r G^{\circ,\ddagger}$ values calculated at the ωB97M-D4/def2-QZVPP//r^{2}SCAN-3c level of theory.

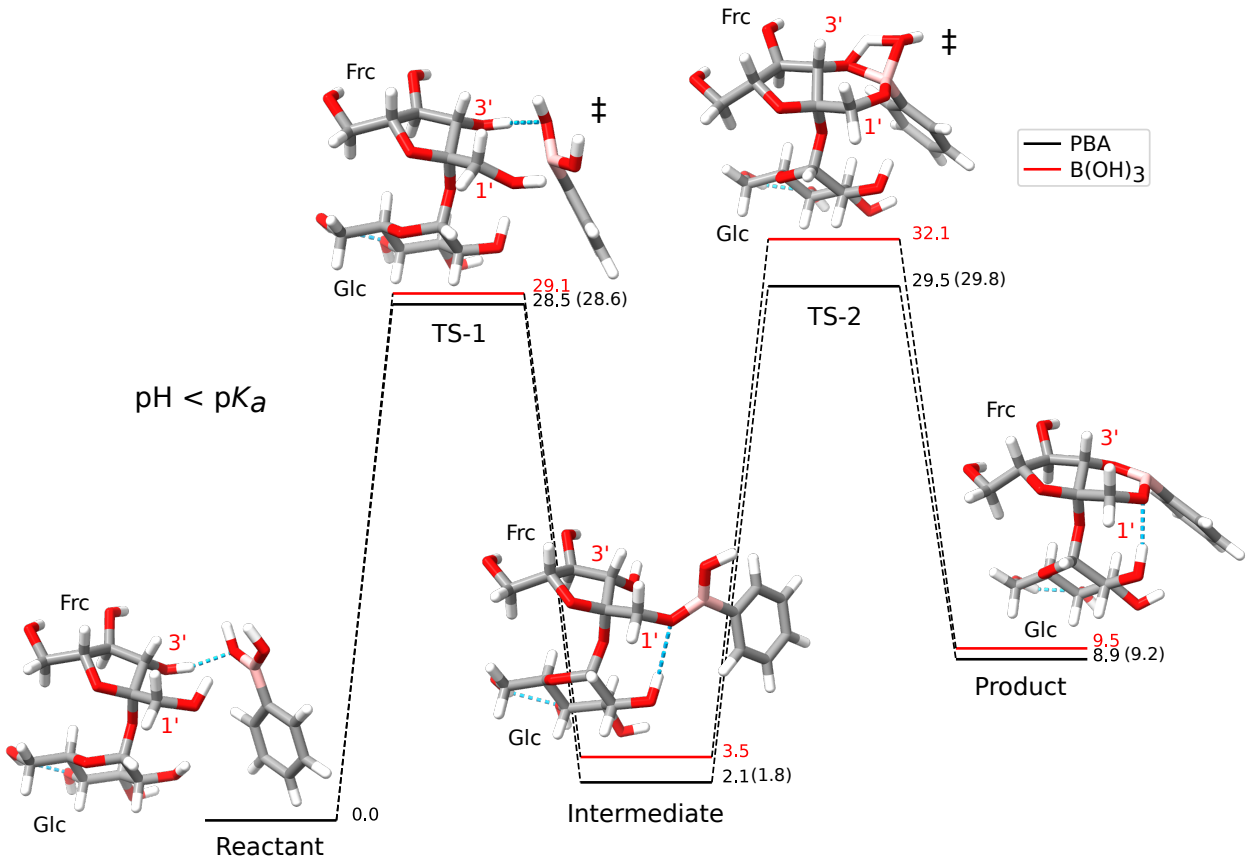


Figure C.1: Same as Figure 4.7 but for PBA (black line) and $B(OH)_3$ (red line) reacting with the fructose moiety of sucrose at reaction conditions pH $< pK_a$ of the boronic compound. For PBA, the imaginary frequencies for transition states TS-1 and TS-2 are $1343.38i$ cm^{-1} and $1374.19i$ cm^{-1}, respectively. For $B(OH)_3$, the corresponding imaginary frequencies are $1383.39i$ cm^{-1} for TS-1 and $1406.31i$ cm^{-1} for TS-2. Dashed blue lines highlight the hydrogen bonds.

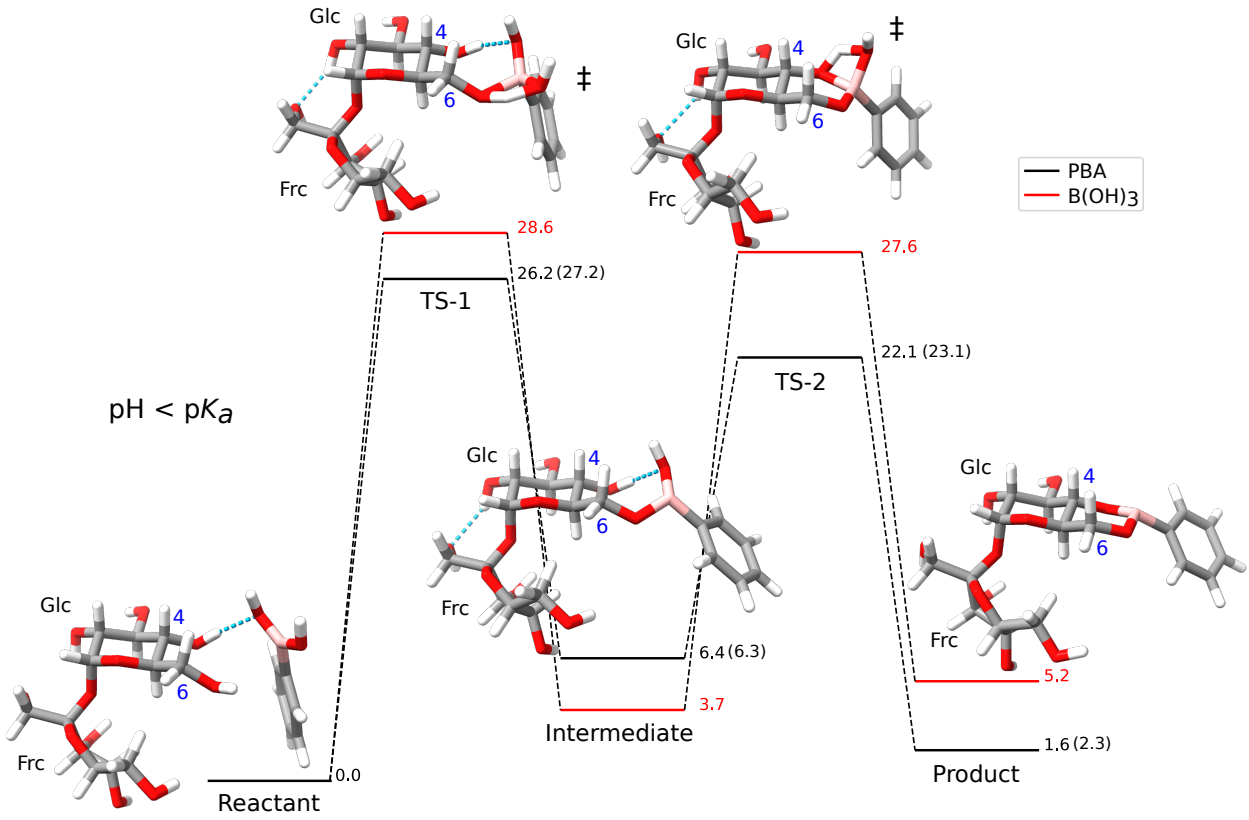


Figure C.2: Same as Figure 4.7 but for PBA (black line) and $B(OH)_3$ (red line) reacting with the glucose moiety of sucrose at reaction conditions pH $<$ pK_a of the boronic compound. For PBA, the imaginary frequencies for transition states TS-1 and TS-2 are $1326.39i$ cm^{-1} and $1392.92i$ cm^{-1}, respectively. For $B(OH)_3$, the corresponding imaginary frequencies are $1406.31i$ cm^{-1} for TS-1 and $1392.06i$ cm^{-1} for TS-2. Dashed blue lines highlight the hydrogen bonds.

Appendix D

NMR Characterization of Compounds for Synthesis of a DBA Tweezer

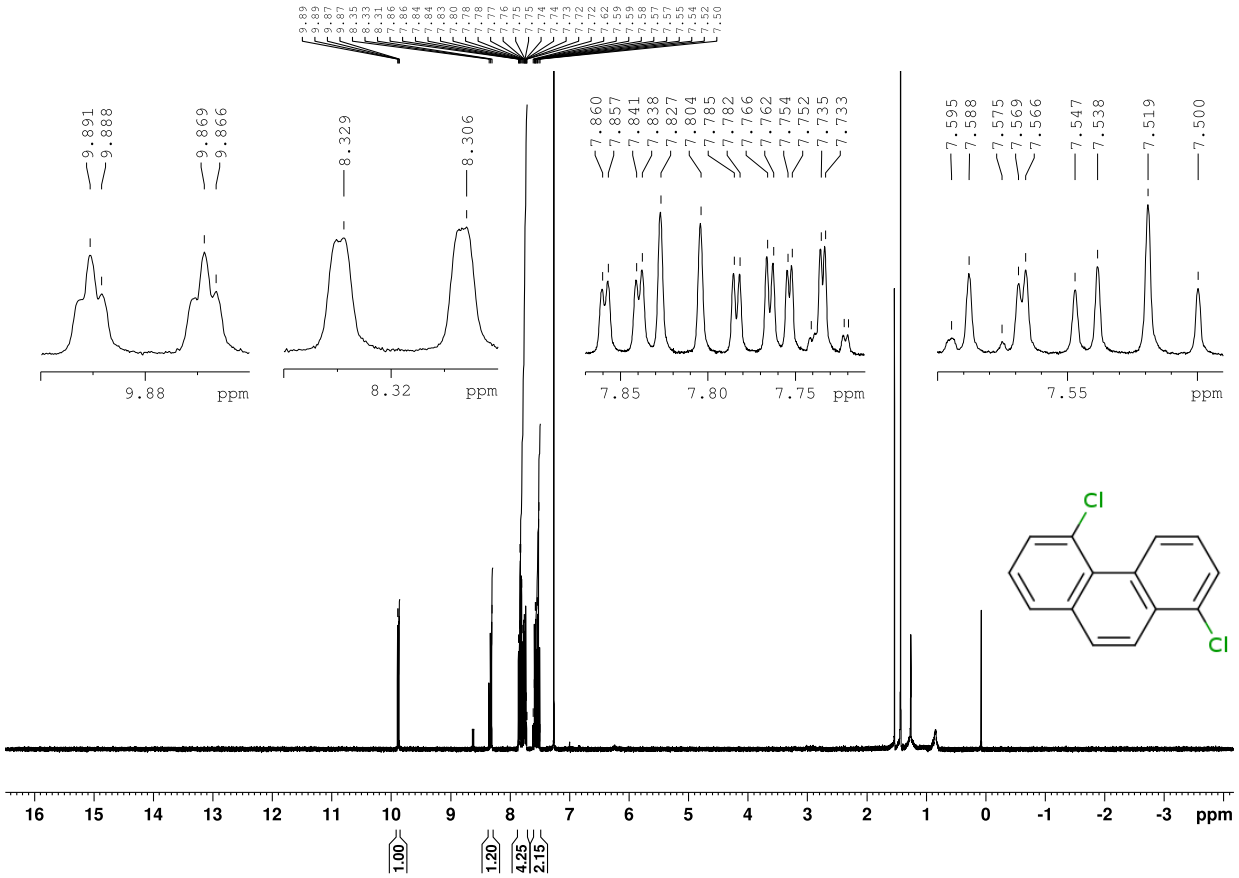


Figure D.1: ^{1}H NMR spectra of compound **P1** (400 MHz, $CDCl_3$, 25 °C)

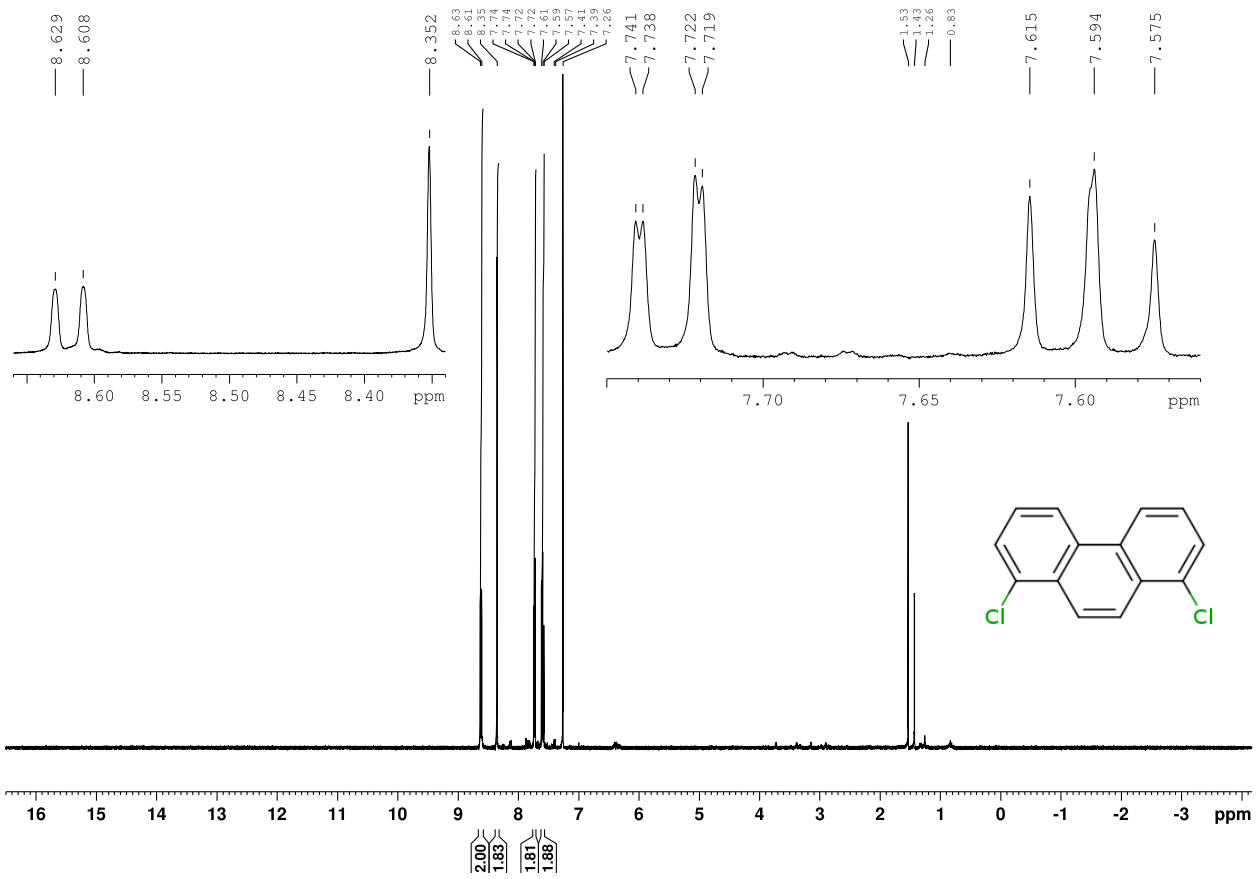


Figure D.2: ^{1}H NMR spectra of compound **P1B** (400 MHz, $CDCl_3$, 25 °C)

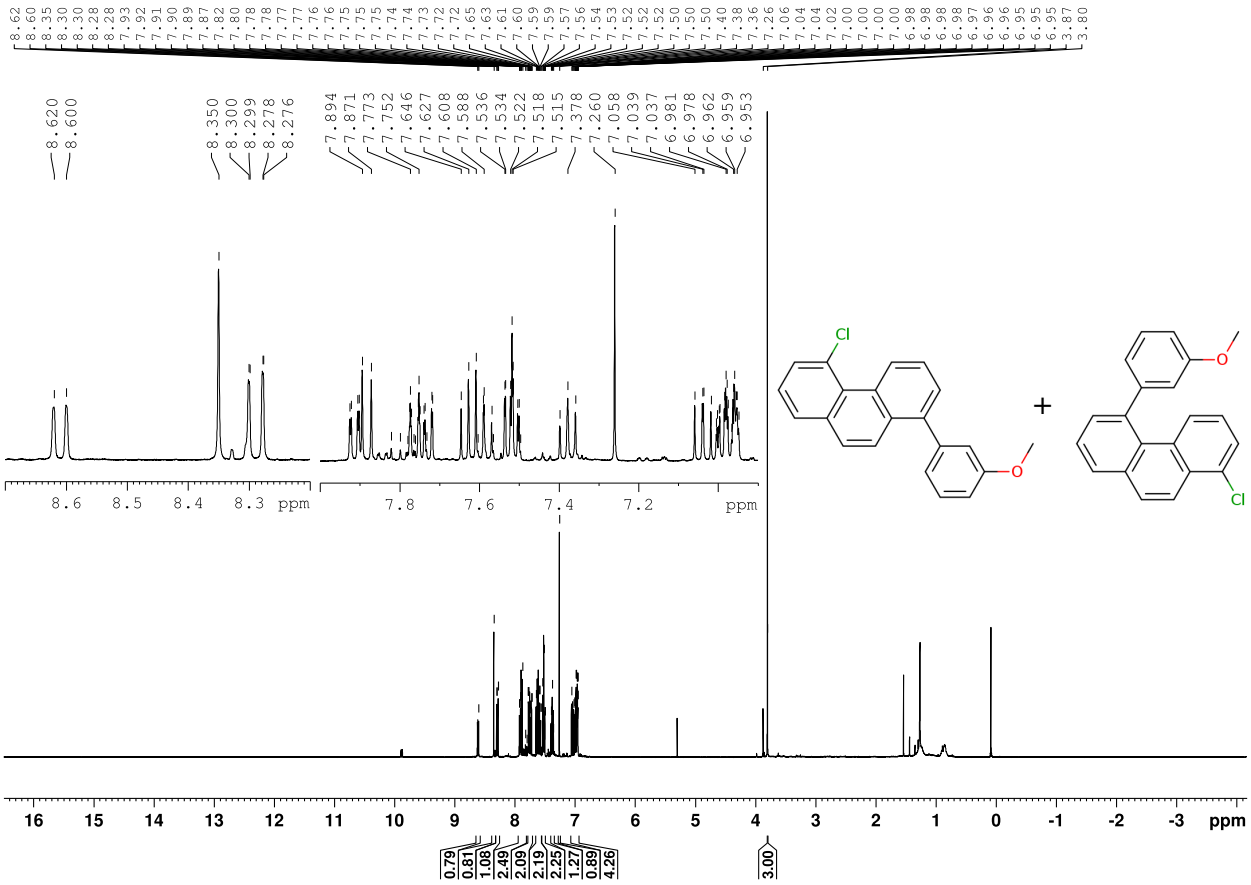


Figure D.3: ^{1}H NMR spectra of compounds **P1.5A** and **P1.5B** (400 MHz, $CDCl_3$, 25 °C)

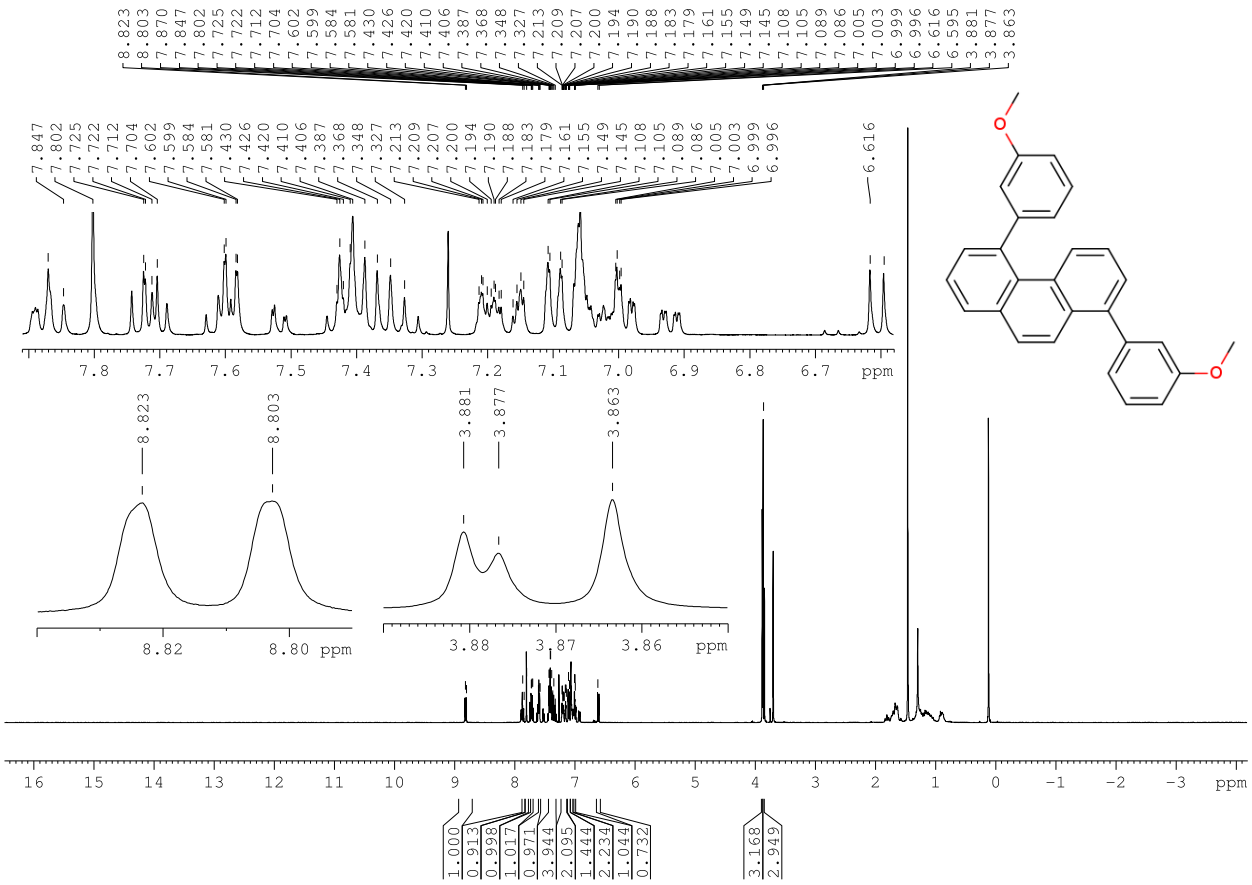


Figure D.4: ^{1}H NMR spectra of compound **P2** (400 MHz, $CDCl_3$, 25 °C)

Bibliography

[1] J. A. Pounds and R. Puschendorf, *Clouded futures*, Nature **427**, 107–109 (2004).

[2] W. R. Cline, *Global Warming and Agriculture*, Finance and Development **0045**, A007 (2008).

[3] S. Chakraborty and A. C. Newton, *Climate change, plant diseases and food security: an overview*, Plant Pathology **60**, 2–14 (2011).

[4] Z. Rengel. *Soil pH, Soil Health and Climate Change*, pages 69–85. Springer Berlin Heidelberg, Berlin, Heidelberg, (2011).

[5] A. Razzaq, P. Kaur, N. Akhter, S. H. Wani, and F. Saleem, *Next-Generation Breeding Strategies for Climate-Ready Crops*, Frontiers in Plant Science **12** (2021).

[6] S. Varotto, T. Krugman, R. Aiese Cigliano, K. Kashkush, A. Kondić-Špika, F. A. Aravanopoulos, M. Pradillo, F. Consiglio, R. Aversano, A. Pecinka, and D. Miladinović, *Exploitation of epigenetic variation of crop wild relatives for crop improvement and agrobiodiversity preservation*, Theoretical and Applied Genetics **135**, 3987–4003 (2022).

[7] N. Halford, T. Curtis, N. Muttucumaru, J. Postles, and D. Mottram, *Sugars in crop plants*, Annals of Applied Biology **158**, 1–25 (2011).

[8] J. Yoon, L.-H. Cho, W. Tun, J.-S. Jeon, and G. An, *Sucrose signaling in higher plants*, Plant Science **302**, 110703 (2021).

[9] I. Ciereszko, *Regulatory roles of sugars in plant growth and development*, Acta Societatis Botanicorum Poloniae **87** (2018).

[10] S. Okumoto, A. Jones, and W. B. Frommer, *Quantitative Imaging with Fluorescent Biosensors*, Annual Review of Plant Biology **63**, 663–706 (2012).

[11] M. Sadoine, M. Reger, K. M. Wong, and W. B. Frommer, *Affinity Series of Genetically Encoded Förster Resonance Energy-Transfer Sensors for Sucrose*, ACS Sens. **6**, 1779–1784 (2021).

[12] G. A. Lara-Cruz and A. Jaramillo-Botero, *Molecular Level Sucrose Quantification: A Critical Review*, Sensors **22**, 9511 (2022).

[13] M. Yamashina, M. Akita, T. Hasegawa, S. Hayashi, and M. Yoshizawa, *A polyaromatic nanocapsule as a sucrose receptor in water*, Science Advances **3**, e1701126 (2017).

[14] N. Nakatsuka, K.-A. Yang, J. M. Abendroth, K. M. Cheung, X. Xu, H. Yang, C. Zhao, B. Zhu, Y. S. Rim, Y. Yang, P. S. Weiss, M. N. Stojanović, and A. M. Andrews, *Aptamer-field-effect transistors overcome Debye length limitations for small-molecule sensing*, Science **362**, 319–324 (2018).

[15] H. Shekarchizadeh, A. A. Ensafi, and M. Kadivar, *Selective determination of sucrose based on electropolymerized molecularly imprinted polymer modified multiwall carbon nanotubes/glassy carbon electrode*, Materials Science and Engineering: C **33**, 3553–3561 (2013).

[16] T. Sakata, S. Nishitani, and T. Kajisa, *Molecularly imprinted polymer-based bioelectrical interfaces with intrinsic molecular charges*, RSC Adv. **10**, 16999–17013 (2020).

[17] W. Yang, H. He, and D. G. Drueckhammer, *Computer-Guided Design in Molecular Recognition: Design and Synthesis of a Glucopyranose Receptor*, Angewandte Chemie International Edition **40**, 1714–1718 (2001).

[18] Z. Bian, A. Liu, Y. Li, G. Fang, Q. Yao, G. Zhang, and Z. Wu, *Boronic acid sensors with double recognition sites: a review*, Analyst **145**, 719–744 (2020).

[19] A. E. Rowan, S. J. Rowan, T. Aida, T. D. James, M. D. Phillips, and S. Shinkai, *Boronic Acids in Saccharide Recognition*, The Royal Society of Chemistry (2006).

[20] T. Kajisa and T. Sakata, *Fundamental Properties of Phenylboronic-Acid-Coated Gate Field-Effect Transistor for Saccharide Sensing*, ChemElectroChem **1**, 1647–1655 (2014).

[21] Y. Egawa, T. Seki, S. Takahashi, and J. ichi Anzai, *Electrochemical and optical sugar sensors based on phenylboronic acid and its derivatives*, Materials Science and Engineering: C **31**, 1257–1264 (2011).

[22] S. Seraj, S. Rouhani, Z. Ranjbar, and S. L. Esfahani, *Fructose recognition using novel solid-state electro-optical nanosensor based on boronate-tagged fluorophore modified graphene oxide*, Materials Chemistry and Physics **270**, 124842 (2021).

[23] E. Shoji and M. S. Freund, *Potentiometric Saccharide Detection Based on the pKa Changes of Poly(aniline boronic acid)*, J. Am. Chem. Soc. **124**, 12486–12493 (2002).

[24] F. Wang, M. Lu, H. Yuan, Y. Zhang, W. Ji, C. Sun, and W. Peng, *pM Level and Large Dynamic Range Glucose Detection Based on a Sandwich Type Plasmonic Fiber Sensor*, Journal of Lightwave Technology **39**, 3882–3889 (2021).

[25] S. Qian, Y. Liang, J. Ma, Y. Zhang, J. Zhao, and W. Peng, *Boronic acid modified fiber optic SPR sensor and its application in saccharide detection*, Sensors and Actuators B: Chemical **220**, 1217–1223 (2015).

[26] H. Yuan, W. Ji, S. Chu, S. Qian, F. Wang, J.-F. Masson, X. Han, and W. Peng, *Fiber-optic surface plasmon resonance glucose sensor enhanced with phenylboronic acid modified Au nanoparticles*, Biosensors and Bioelectronics **117**, 637–643 (2018).

[27] K. V. Kong, Z. Lam, W. K. O. Lau, W. K. Leong, and M. Olivo, *A Transition Metal Carbonyl Probe for Use in a Highly Specific and Sensitive SERS-Based Assay for Glucose*, Journal of the American Chemical Society **135**, 18028–18031 (2013).

[28] M. Pushina, A. Penavic, S. Farshbaf, and P. Anzenbacher, *Fluorescent Sensor Array for Quantitative Determination of Saccharides*, ACS Sens. **6**, 4001–4008 (2021).

[29] F. D'Hooge, S. A. Elfeky, S. E. Flower, S. I. Pascu, A. T. A. Jenkins, J. M. H. v. d. Elsen, T. D. James, and J. S. Fossey, *Biotinylated boronic acid fluorophore conjugates: Quencher elimination strategy for imaging and saccharide detection*, RSC Adv. **2**, 3274–3280 (2012).

[30] K. R. A. S. Sandanayake, K. Nakashima, and S. Shinkai, *Specific recognition of disaccharides by trans-3,3'-stilbenediboronic acid: rigidification and fluoresecence enhancement of the stilbene skeleton upon formation of a sugar–stilbene macrocycle*, J. Chem. Soc., Chem. Commun. –, 1621–1622 (1994).

[31] X.-t. Zhang, S. Wang, and G.-w. Xing, *Novel Boronlectins Based on Bispyridium Salt with a Flexible Linker: Discriminative Sensing of Lactose and Other Monosaccharides*

and Disaccharides in Aqueous Solution, Chemistry – An Asian Journal **10**, 2594–2598 (2015).

[32] L. Chi, J. Zhao, and T. D. James, *Chiral Mono Boronic Acid As Fluorescent Enantioselective Sensor for Mono α-Hydroxyl Carboxylic Acids*, J. Org. Chem. **73**, 4684–4687 (2008).

[33] G. T. Williams, J. L. Kedge, and J. S. Fossey, *Molecular Boronic Acid-Based Saccharide Sensors*, ACS Sens. **6**, 1508–1528 (2021).

[34] G. M. Brown and H. A. Levy, *Further refinement of the structure of sucrose based on neutron-diffraction data*, Acta Crystallographica Section B **29**, 790–797 (1973).

[35] T. d. C. Rozada, R. M. Pontes, R. Rittner, and E. A. Basso, *Stereoelectronic effects of the glycosidic linkage on the conformational preference of d-sucrose*, RSC Adv. **6**, 112806–112812 (2016).

[36] R. T. Morrison and R. N. Boyd, *Organic chemistry*, Allyn and Bacon Boston, 5h ed. edition (1973).

[37] H. Manoochehri, N. F. Hosseini, M. Saidijam, M. Taheri, H. Rezaee, and F. Nouri, *A review on invertase: Its potentials and applications*, Biocatalysis and Agricultural Biotechnology **25**, 101599 (2020).

[38] O. Stein and D. Granot, *An Overview of Sucrose Synthases in Plants*, Frontiers in Plant Science **10**, 95 (2019).

[39] J. D. Eastin, F. A. Haskins, C. Y. Sullivan, C. H. M. V. Bavel, and R. C. Dinauer, editors. *Mechanisms of Translocation of Plant Metabolites*, Madison, Wisconsin, (1979). American Society of Agronomy & Crop Science Society of America, UNL Digital Commons. Published in Physiological Aspects of Crop Yield: Proceedings of a symposium sponsored by the University of Nebraska, the American Society of Agronomy, and the Crop Science Society of America, and held at the University of Nebraska, Lincoln, Nebr., January 20-24, 1969.

[40] C. Diacci, T. Abedi, J. W. Lee, E. O. Gabrielsson, M. Berggren, D. T. Simon, T. Niittylä, and E. Stavrinidou, *Diurnal in vivo xylem sap glucose and sucrose monitoring using implantable organic electrochemical transistor sensors*, iScience **24**, 101966 (2021).

[41] S. A. Perdomo, E. De la Paz, R. Del Caño, S. Seker, T. Saha, J. Wang, and A. Jaramillo-Botero, *Non-invasive in-vivo glucose-based stress monitoring in plants*, Biosensors and Bioelectronics **231**, 115300 (2023).

[42] H. Shibata, Y. J. Heo, T. Okitsu, Y. Matsunaga, T. Kawanishi, and S. Takeuchi, *Injectable hydrogel microbeads for fluorescence-based in vivo continuous glucose monitoring*, Proceedings of the National Academy of Sciences **107**, 17894–17898 (2010).

[43] K. Wang, R. Zhang, X. Zhao, Y. Ma, L. Ren, Y. Ren, G. Chen, D. Ye, J. Wu, X. Hu, Y. Guo, R. Xi, M. Meng, Q. Yao, P. Li, Q. Chen, and T. D. James, *Reversible Recognition-Based Boronic Acid Probes for Glucose Detection in Live Cells and Zebrafish*, Journal of the American Chemical Society **145**, 8408–8416 (2023).

[44] K. Wang, R. Zhang, X. Yue, Z. Zhou, L. Bai, Y. Tong, B. Wang, D. Gu, S. Wang, Y. Qiao, Q. Liu, X. Xue, Y. Yin, R. Xi, and M. Meng, *Synthesis of Diboronic Acid-Based Fluorescent Probes for the Sensitive Detection of Glucose in Aqueous Media and Biological Matrices*, ACS Sensors **6**, 1543–1551 (2021).

[45] X. Wu, Z. Li, X.-X. Chen, J. S. Fossey, T. D. James, and Y.-B. Jiang, *Selective sensing of saccharides using simple boronic acids and their aggregates*, Chem. Soc. Rev. **42**, 8032–8048 (2013).

[46] Y. Furikado, T. Nagahata, T. Okamoto, T. Sugaya, S. Iwatsuki, M. Inamo, H. D. Takagi, A. Odani, and K. Ishihara, *Universal Reaction Mechanism of Boronic Acids with Diols in Aqueous Solution: Kinetics and the Basic Concept of a Conditional Formation Constant*, Chemistry – A European Journal **20**, 13194–13202 (2014).

[47] Z. Liu and H. He, *Synthesis and Applications of Boronate Affinity Materials: From Class Selectivity to Biomimetic Specificity*, Acc. Chem. Res. **50**, 2185–2193 (2017).

[48] J. A. Peters, *Interactions between boric acid derivatives and saccharides in aqueous media: Structures and stabilities of resulting esters*, Coordination Chemistry Reviews **268**, 1–22 (2014).

[49] J. Yan, G. Springsteen, S. Deeter, and B. Wang, *The relationship among pKa, pH, and binding constants in the interactions between boronic acids and diols—it is not as simple as it appears*, Tetrahedron **60**, 11205–11209 (2004).

[50] Y. Ohno, R. Tanaka, Y. Suzuki, T. Sugaya, S. Iwatsuki, M. Inamo, and K. Ishihara, *Detailed Reaction Mechanism of Bis-(o-Aminomethylphenylboronic Acid)-based Receptors with Various Length Methylene-chain Linkers with D-Glucose*, ChemistrySelect **7** (2022).

[51] M. J. O'Neil, *The Merck index : an encyclopedia of chemicals, drugs, and biologicals*, Merck, Whitehouse Station, N.J. (2001).

[52] Y. Suzuki, M. Shimizu, T. Okamoto, T. Sugaya, S. Iwatsuki, M. Inamo, H. D. Takagi, A. Odani, and K. Ishihara, *Detailed Mechanism of the Reaction of Phenylboronic Acid Derivatives with D-Fructose in Aqueous Solution: A Comprehensive Kinetic Study*, ChemistrySelect **1**, 5141–5151 (2016).

[53] K. Tsukagoshi and S. Shinkai, *Specific complexation with mono- and disaccharides that can be detected by circular dichroism*, The Journal of Organic Chemistry **56**, 4089–4091 (1991).

[54] T. D. James, K. R. A. S. Sandanayake, and S. Shinkai, *A Glucose-Selective Molecular Fluorescence Sensor*, Angewandte Chemie International Edition in English **33**, 2207–2209 (1994).

[55] H. Eggert, J. Frederiksen, C. Morin, and J. C. Norrild, *A New Glucose-Selective Fluorescent Bisboronic Acid. First Report of Strong α-Furanose Complexation in Aqueous Solution at Physiological pH1*, The Journal of Organic Chemistry **64**, 3846–3852 (1999).

[56] S. Arimori, M. L. Bell, C. S. Oh, K. A. Frimat, and T. D. James, *Modular fluorescence sensors for saccharides*, Chem. Commun. **37**, 1836–1837 (2001).

[57] Q. Chen, Y. Fu, W. Zhang, S. Ye, H. Zhang, F. Xie, L. Gong, Z. Wei, H. Jin, and J. Chen, *Highly sensitive detection of glucose: A quantitative approach employing nanorods assembled plasmonic substrate*, Talanta **165**, 516–521 (2017).

[58] B. Wang, K.-H. Chou, B. N. Queenan, S. Pennathur, and G. C. Bazan, *Molecular Design of a New Diboronic Acid for the Electrohydrodynamic Monitoring of Glucose*, Angewandte Chemie International Edition **58**, 10612–10615 (2019).

[59] J. D. Larkin, K. A. Frimat, T. M. Fyles, S. E. Flower, and T. D. James, *Boronic acid based photoinduced electron transfer (PET) fluorescence sensors for saccharides*, New J. Chem. **34**, 2922–2931 (2010).

[60] B. P. Hay and T. K. Firman, *HostDesigner: A Program for the de Novo Structure-Based Design of Molecular Receptors with Binding Sites that Complement Metal Ion Guests*, Inorganic Chemistry **41**, 5502–5512 (2002).

[61] B. P. Hay, C. Jia, and J. Nadas, *Computer-aided design of host molecules for recognition of organic guests*, Computational and Theoretical Chemistry **1028**, 72–80 (2014).

[62] M. Foscato, V. Venkatraman, and V. R. Jensen, *DENOPTIM: Software for Computational de Novo Design of Organic and Inorganic Molecules*, Journal of Chemical Information and Modeling **59**, 4077–4082 (2019).

[63] A. Thakkar, V. Chadimová, E. J. Bjerrum, O. Engkvist, and J.-L. Reymond, *Retrosynthetic accessibility score (RAscore) – rapid machine learned synthesizability classification from AI driven retrosynthetic planning*, Chem. Sci. **12**, 3339–3349 (2021).

[64] P. Schwaller, R. Petraglia, V. Zullo, V. H. Nair, R. A. Haeuselmann, R. Pisoni, C. Bekas, A. Iuliano, and T. Laino, *Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy*, Chem. Sci. **11**, 3316–3325 (2020).

[65] A. Toniato, P. Schwaller, A. Cardinale, J. Geluykens, and T. Laino, *Unassisted noise reduction of chemical reaction datasets*, Nature Machine Intelligence **3**, 485–494 (2021).

[66] J. H. Jensen, *A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space*, Chemical Science **10**, 3567–3572 (2019).

[67] T. Baer and W. L. Hase, *Unimolecular reaction dynamics: theory and experiments*, Oxford University Press, New York (1996).

[68] B. T. Sutcliffe, *The idea of a potential energy surface*, Molecular Physics **104**, 715–722 (2006).

[69] I. N. Levine, *Química Cuántica*, Pearson Educación (2001).

[70] J. S. Tse, *AB INITIO MOLECULAR DYNAMICS WITH DENSITY FUNCTIONAL THEORY*, Annual Review of Physical Chemistry **53**, 249–290 (2002).

[71] W. F. van Gunsteren and H. J. C. Berendsen, *Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry*, Angewandte Chemie International Edition in English **29**, 992–1023 (1990).

[72] P. Hohenberg and W. Kohn, *Inhomogeneous Electron Gas*, Phys. Rev. **136**, B864–B871 (1964).

[73] W. Kohn and L. J. Sham, *Self-Consistent Equations Including Exchange and Correlation Effects*, Phys. Rev. **140**, A1133–A1138 (1965).

[74] E. Runge and E. K. U. Gross, *Density-Functional Theory for Time-Dependent Systems*, Phys. Rev. Lett. **52**, 997–1000 (1984).

[75] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, Wiley, 2 edition (2004).

[76] F. Jensen, *Introduction to Computational Chemistry*, Wiley, 3rd edition (2017).

[77] K. Chenoweth, A. C. T. van Duin, and W. A. Goddard, *ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation*, The Journal of Physical Chemistry A **112**, 1040–1053 (2008).

[78] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, *Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions*, The Journal of Physical Chemistry B **113**, 6378–6396 (2009).

[79] J. M. Herbert, *Dielectric continuum methods for quantum chemistry*, WIREs Computational Molecular Science **11** (2021).

[80] V. S. Bryantsev, M. S. Diallo, and W. A. Goddard III, *Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum Models*, The Journal of Physical Chemistry B **112**, 9709–9719 (2008).

[81] J. H. Jensen, *Predicting accurate absolute binding energies in aqueous solution: thermodynamic considerations for electronic structure methods*, Phys. Chem. Chem. Phys. **17**, 12441–12451 (2015).

[82] J. R. Pliego and J. M. Riveros, *The Cluster-Continuum Model for the Calculation of the Solvation Free Energy of Ionic Species*, The Journal of Physical Chemistry A **105**, 7241–7247 (2001).

[83] J. R. Pliego Jr and J. M. Riveros, *Hybrid discrete-continuum solvation methods*, WIREs Computational Molecular Science **10**, e1440 (2020).

[84] S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher, and M. Stahn, *Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules*, J. Phys. Chem. A **125**, 4039–4054 (2021).

[85] A. Viayna, S. Pinheiro, C. Curutchet, F. J. Luque, and W. J. Zamora, *Prediction of n-octanol/water partition coefficients and acidity constants (pKa) in the SAMPL7 blind challenge with the IEFPCM-MST model*, Journal of Computer-Aided Molecular Design **35**, 803–811 (2021).

[86] J. A. Pople, *Nobel Lecture: Quantum chemical models*, Rev. Mod. Phys. **71**, 1267–1274 (1999).

[87] K. A. Peterson, D. Feller, and D. A. Dixon, *Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges*, Theoretical Chemistry Accounts **131** (2012).

[88] A. Raghav, R. Maezono, K. Hongo, S. Sorella, and K. Nakano, *Toward Chemical Accuracy Using the Jastrow Correlated Antisymmetrized Geminal Power Ansatz*, Journal of Chemical Theory and Computation **19**, 2222–2229 (2023).

[89] A. Karton, *A computational chemist's guide to accurate thermochemistry for organic molecules*, WIREs Computational Molecular Science **6**, 292–310 (2016).

[90] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme, *Best-Practice DFT Protocols for Basic Molecular Computational Chemistry***, Angewandte Chemie International Edition **61**, e202205735 (2022).

[91] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, *Quantum chemical accuracy from density functional approximations via machine learning*, Nature Communications **11** (2020).

[92] E. O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, and A. Aspuru-Guzik, *What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery*, Annual Review of Materials Research **45**, 195–216 (2015).

[93] J. Meyers, B. Fabian, and N. Brown, *De novo molecular design and generative models*, Drug Discovery Today **26**, 2707–2715 (2021).

[94] H. Park, S. Majumdar, X. Zhang, J. Kim, and B. Smit, *Inverse design of metal–organic frameworks for direct air capture of CO2via deep reinforcement learning*, Digital Discovery **3**, 728–741 (2024).

[95] N. Brown, B. McKay, F. Gilardoni, and J. Gasteiger, *A Graph-Based Genetic Algorithm and Its Application to the Multiobjective Evolution of Median Molecules*, Journal of Chemical Information and Computer Sciences **44**, 1079–1087 (2004).

[96] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, *Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems*, Chemical Reviews **121**, 9816–9872 (2021).

[97] Y.-F. Shi, Z.-X. Yang, S. Ma, P.-L. Kang, C. Shang, P. Hu, and Z.-P. Liu, *Machine Learning for Chemistry: Basics and Applications*, Engineering **27**, 70–83 (2023).

[98] N. Yoshikawa, K. Terayama, M. Sumita, T. Homma, K. Oono, and K. Tsuda, *Population-based De Novo Molecule Generation, Using Grammatical Evolution*, Chemistry Letters **47**, 1431–1434 (2018).

[99] N. Brown, M. Fiscato, M. H. Segler, and A. C. Vaucher, *GuacaMol: Benchmarking Models for de Novo Molecular Design*, Journal of Chemical Information and Modeling **59**, 1096–1108 (2019).

[100] P. A. Vikhar. *Evolutionary algorithms: A critical review and its future prospects*. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, pages 261–265, (2016).

[101] J. F. Miller, *Cartesian Genetic Programming*, Springer Berlin Heidelberg (2011).

[102] Y. Kwon and J. Lee, *MolFinder: an evolutionary algorithm for the global optimization of molecular properties and the extensive exploration of chemical space using SMILES*, Journal of Cheminformatics **13** (2021).

[103] A. M. Virshup, J. Contreras-García, P. Wipf, W. Yang, and D. N. Beratan, *Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds*, Journal of the American Chemical Society **135**, 7296–7303 (2013).

[104] M. Krenn, Q. Ai, S. Barthel, N. Carson, A. Frei, N. C. Frey, P. Friederich, T. Gaudin, A. A. Gayle, K. M. Jablonka, R. F. Lameiro, D. Lemm, A. Lo, S. M. Moosavi, J. M. Nápoles-Duarte, A. Nigam, R. Pollice, K. Rajan, U. Schatzschneider, P. Schwaller, M. Skreta, B. Smit, F. Strieth-Kalthoff, C. Sun, G. Tom, G. Falk von Rudorff, A. Wang, A. D. White, A. Young, R. Yu, and A. Aspuru-Guzik, *SELFIES and the future of molecular string representations*, Patterns **3**, 100588 (2022).

[105] D. Weininger, *SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules*, Journal of Chemical Information and Computer Sciences **28**, 31–36 (1988).

[106] RDKit: Open-source cheminformatics. https://www.rdkit.org.

[107] I. Daylight Chemical Information Systems. *SMARTS - A Language for Describing Molecular Patterns*, (2019).

[108] G. A. Lara-Cruz, T. Rose, S. Grimme, and A. Jaramillo-Botero, *Reaction-Free Energies for Complexation of Carbohydrates by Tweezer Diboronic Acids*, The Journal of Physical Chemistry B **128**, 9213–9223 (2024).

[109] J. Xia and D. A. Case, *Sucrose in aqueous solution revisited, Part 1: Molecular dynamics simulations and direct and indirect dipolar coupling analysis*, Biopolymers **97**, 276–288 (2012).

[110] P. Pracht, F. Bohle, and S. Grimme, *Automated exploration of the low-energy chemical space with fast quantum chemical methods*, Phys. Chem. Chem. Phys. **22**, 7169–7192 (2020).

[111] S. Spicher, C. Plett, P. Pracht, A. Hansen, and S. Grimme, *Automated Molecular Cluster Growing for Explicit Solvation by Efficient Force Field and Tight Binding Methods*, J. Chem. Theory Comput. **18**, 3174–3189 (2022).

[112] S. Spicher and S. Grimme, *Single-Point Hessian Calculations for Improved Vibrational Frequencies and Rigid-Rotor-Harmonic-Oscillator Thermodynamics*, Journal of Chemical Theory and Computation **17**, 1701–1714 (2021).

[113] S. Ehlert, M. Stahn, S. Spicher, and S. Grimme, *Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods*, Journal of Chemical Theory and Computation **17**, 4250–4261 (2021).

[114] A. Onufriev, D. Bashford, and D. A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized born model*, Proteins: Struct., Funct., Bioinf. **55**, 383–394 (2004).

[115] G. Sigalov, A. Fenley, and A. Onufriev, *Analytical electrostatics for biomolecules: Beyond the generalized Born approximation*, J. Chem. Phys. **124**, 124902 (2006).

[116] A. W. Lange and J. M. Herbert, *Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. I. An Improved Effective Coulomb Operator*, J. Chem. Theory Comput. **8**, 1999–2011 (2012).

[117] J. C. Norrild and H. Eggert, *Evidence for Mono- and Bisdentate Boronate Complexes of Glucose in the Furanose Form. Application of 1JC-C Coupling Constants as a Structural Probe*, Journal of the American Chemical Society **117**, 1479–1484 (1995).

[118] F. Urban and P. A. Shaffer, *THE ACIDIC PROPERTY OF SUGARS*, Journal of Biological Chemistry **94**, 697–715 (1932).

[119] F. Urban and R. Williams, *THE ACIDIC PROPERTY OF SUGARS. II*, Journal of Biological Chemistry **100**, 237–241 (1933).

[120] Y. Shiomi, M. Saisho, K. Tsukagoshi, and S. Shinkai, *Specific complexation of glucose with a diphenylmethane-3,3′-diboronic acid derivative: correlation between the absolute configuration of mono- and di-saccharides and the circular dichroic activity of the complex*, J. Chem. Soc., Perkin Trans. 1 , 2111–2117 (1993).

[121] J. G. Brandenburg, T. Maas, and S. Grimme, *Benchmarking DFT and semiempirical methods on structures and lattice energies for ten ice polymorphs*, J. Chem. Phys. **142**, 124104 (2015).

[122] S. Ehlert, S. Grimme, and A. Hansen, *Conformational Energy Benchmark for Longer n-Alkane Chains*, J. Phys. Chem. A **126**, 3521–3535 (2022).

[123] M. Müller, A. Hansen, and S. Grimme, *ωB97X-3c: A composite range-separated hybrid DFT method with a molecule-optimized polarized valence double-ζ basis set*, J. Chem. Phys. **158**, 014103 (2023).

[124] S. Spicher, E. Caldeweyher, A. Hansen, and S. Grimme, *Benchmarking London dispersion corrected density functional theory for noncovalent ion–π interactions*, Phys. Chem. Chem. Phys. **23**, 11635–11648 (2021).

[125] S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen, *Consistent structures and interactions by density functional theory with small atomic orbital basis sets*, The Journal of Chemical Physics **143**, 054107 (2015).

[126] F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, *The ORCA quantum chemistry program package*, The Journal of Chemical Physics **152**, 224108 (2020).

[127] G. Mills, H. Jónsson, and G. K. Schenter, *Reversible work transition state theory: application to dissociative adsorption of hydrogen*, Surface Science **324**, 305–337 (1995).

[128] V. Ásgeirsson, B. O. Birgisson, R. Bjornsson, U. Becker, F. Neese, C. Riplinger, and H. Jónsson, *Nudged Elastic Band Method for Molecular Reactions Using Energy-Weighted Springs Combined with Eigenvector Following*, Journal of Chemical Theory and Computation **17**, 4929–4945 (2021).

[129] S. Smidstrup, A. Pedersen, K. Stokbro, and H. Jónsson, *Improved initial guess for minimum energy path calculations*, The Journal of Chemical Physics **140** (2014).

[130] S. Grimme, A. Hansen, S. Ehlert, and J.-M. Mewes, *r2SCAN-3c: A "Swiss army knife" composite electronic-structure method*, The Journal of Chemical Physics **154** (2021).

[131] V. Barone and M. Cossi, *Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model*, The Journal of Physical Chemistry A **102**, 1995–2001 (1998).

[132] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, *Open Babel: An open chemical toolbox*, Journal of Cheminformatics **3** (2011).

[133] C. Bannwarth, S. Ehlert, and S. Grimme, *GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions*, Journal of Chemical Theory and Computation **15**, 1652–1671 (2019).

[134] X. Pan, H. Wang, C. Li, J. Z. H. Zhang, and C. Ji, *MolGpka: A Web Server for Small Molecule pKa Prediction Using a Graph-Convolutional Neural Network*, Journal of Chemical Information and Modeling **61**, 3159–3165 (2021).

[135] D. Rogers and M. Hahn, *Extended-Connectivity Fingerprints*, Journal of Chemical Information and Modeling **50**, 742–754 (2010).

[136] R. Tanaka, K. Matsuo, S. Suzuki, G. Mikami, Y. Suzuki, T. Sugaya, S. Iwatsuki, M. Inamo, and K. Ishihara, *Detailed Reaction Mechanism of Anthracene-Bridged bis-(3-Pyridiniumboronic Acid) Chemosensor with D-Glucose*, ChemistrySelect **9** (2024).

[137] J. W. Raymond and P. Willett, *Effectiveness of graph-based and fingerprint-based similarity measures for virtual screening of 2D chemical structure databases*, Journal of Computer-Aided Molecular Design **16**, 59–71 (2002).

[138] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme, *Extended tight-binding quantum chemistry methods*, WIREs Computational Molecular Science **11** (2020).

[139] P. Ertl and A. Schuffenhauer, *Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions*, Journal of Cheminformatics **1** (2009).

[140] T. Blickle and L. Thiele. *A Comparison of Selection Schemes used in Genetic Algorithms.* (1995).

[141] G. R. Harvey, *Advances in the Synthesis of Polycyclic Aromatic Compounds*, Current Organic Chemistry **8**, 303–323 (2004).

[142] T. Jin and M. Terada, *Recent topics on synthesis of pi-extended polycycles by cascade annulations*, Tetrahedron Letters **61**, 151514 (2020).

[143] E. Pensa, A. A. Rubert, G. Benitez, P. Carro, A. G. Orive, A. H. Creus, R. C. Salvarezza, and C. Vericat, *Are 4-Mercaptobenzoic Acid Self Assembled Monolayers on Au(111) a Suitable System to Test Adatom Models?*, The Journal of Physical Chemistry C **116**, 25765–25771 (2012).

[144] D. Barriet, C. M. Yam, O. E. Shmakova, A. C. Jamison, and T. R. Lee, *4-Mercaptophenylboronic Acid SAMs on Gold: Comparison with SAMs Derived from Thiophenol, 4-Mercaptophenol, and 4-Mercaptobenzoic Acid*, Langmuir **23**, 8866–8875 (2007).

[145] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, *Avogadro: an advanced semantic chemical editor, visualization, and analysis platform*, Journal of Cheminformatics **4** (2012).

[146] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, *LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales*, Comp. Phys. Comm. **271**, 108171 (2022).

[147] S. Jo, T. Kim, V. G. Iyer, and W. Im, *CHARMM-GUI: A web-based graphical user interface for CHARMM*, Journal of Computational Chemistry **29**, 1859–1865 (2008).

[148] S. Jo, X. Cheng, S. M. Islam, L. Huang, H. Rui, A. Zhu, H. S. Lee, Y. Qi, W. Han, K. Vanommeslaeghe, A. D. MacKerell, B. Roux, and W. Im. *CHARMM-GUI PDB Manipulator for Advanced Modeling and Simulations of Proteins Containing Nonstandard Residues*, pages 235–265. Elsevier, (2014).

[149] S.-J. Park, N. Kern, T. Brown, J. Lee, and W. Im, *CHARMM-GUI PDB Manipulator: Various PDB Structural Modifications for Biomolecular Modeling and Simulation*, Journal of Molecular Biology **435**, 167995 (2023).

[150] L. Kong, S.-J. Park, and W. Im, *CHARMM-GUI PDB Reader and Manipulator: Covalent Ligand Modeling and Simulation*, Journal of Molecular Biology **436**, 168554 (2024).

[151] N. R. Kern, J. Lee, Y. K. Choi, and W. Im, *CHARMM-GUI Multicomponent Assembler for modeling and simulation of complex multicomponent systems*, Nature Communications **15** (2024).

[152] K. Vanommeslaeghe and A. D. MacKerell, *Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing*, Journal of Chemical Information and Modeling **52**, 3144–3154 (2012).

[153] J. R. Pliego Jr and J. M. Riveros, *Hybrid discrete-continuum solvation methods*, WIREs Computational Molecular Science **10**, e1440 (2020).

[154] R. Sure, M. el Mahdali, A. Plajer, and P. Deglmann, *Towards a converged strategy for including microsolvation in reaction mechanism calculations*, Journal of Computer-Aided Molecular Design **35**, 473–492 (2021).

[155] G. N. Simm, P. L. Türtscher, and M. Reiher, *Systematic microsolvation approach with a cluster-continuum scheme and conformational sampling*, Journal of Computational Chemistry **41**, 1144–1155 (2020).

[156] J. Zhang, H. Zhang, T. Wu, Q. Wang, and D. van der Spoel, *Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents*, Journal of Chemical Theory and Computation **13**, 1034–1043 (2017).

[157] L. C. Kröger, S. Müller, I. Smirnova, and K. Leonhard, *Prediction of Solvation Free Energies of Ionic Solutes in Neutral Solvents*, J. Phys. Chem. A **124**, 4171–4181 (2020).

[158] C. Plett, M. Stahn, M. Bursch, J.-M. Mewes, and S. Grimme, *Improving Quantum Chemical Solvation Models by Dynamic Radii Adjustment for Continuum Solvation (DRACO)*, J. Phys. Chem. Lett. **15**, 2462–2469 (2024).

[159] M. Stahn, S. Ehlert, and S. Grimme, *Extended Conductor-like Polarizable Continuum Solvation Model (CPCM-X) for Semiempirical Methods*, J. Phys. Chem. A **127**, 7036–7043 (2023).

[160] S. Feng, C. Bagia, and G. Mpourmpakis, *Determination of Proton Affinities and Acidity Constants of Sugars*, The Journal of Physical Chemistry A **117**, 5211–5219 (2013).

[161] W. H. Kruskal and W. A. Wallis, *Use of Ranks in One-Criterion Variance Analysis*, J. Am. Stat. Assoc. **47**, 583–621 (1952).

[162] M. Müller, A. Hansen, and S. Grimme, *An atom-in-molecule adaptive polarized valence single-ζ atomic orbital basis for electronic structure calculations*, The Journal of Chemical Physics **159** (2023).

[163] J. C. Faver, W. Yang, and K. M. Merz, *The Effects of Computational Modeling Errors on the Estimation of Statistical Mechanical Variables*, Journal of Chemical Theory and Computation **8**, 3769–3776 (2012).

[164] Y. Li and B. Hartke, *Assessing Solvation Effects on Chemical Reactions with Globally Optimized Solvent Clusters*, ChemPhysChem **14**, 2678–2686 (2013).

[165] S. Jiang, J. O. Escobedo, K. K. Kim, O. Alptürk, G. K. Samoei, S. O. Fakayode, I. M. Warner, O. Rusin, and R. M. Strongin, *Stereochemical and Regiochemical Trends in the Selective Detection of Saccharides*, J. Am. Chem. Soc. **128**, 12221–12228 (2006).

[166] R. van den Berg, J. A. Peters, and H. van Bekkum, *The structure and (local) stability constants of borate esters of mono- and di-saccharides as studied by 11B and 13C NMR spectroscopy*, Carbohydrate Research **253**, 1–12 (1994).

[167] J. Guo, F. Knuth, C. Margreitter, J. P. Janet, K. Papadopoulos, O. Engkvist, and A. Patronov, *Link-INVENT: generative linker design with reinforcement learning*, Digital Discovery **2**, 392–408 (2023).

[168] C. W. Coley, L. Rogers, W. H. Green, and K. F. Jensen, *SCScore: Synthetic Complexity Learned from a Reaction Corpus*, Journal of Chemical Information and Modeling **58**, 252–261 (2018).

[169] B. Adams and L. Lerner, *Observation of hydroxyl protons of sucrose in aqueous solution: no evidence for persistent intramolecular hydrogen bonds*, Journal of the American Chemical Society **114**, 4827–4829 (1992).

[170] M. D. Battistel, H. F. Azurmendi, and D. I. Freedberg, *Glycan OH Exchange Rate Determination in Aqueous Solution: Seeking Evidence for Transient Hydrogen Bonds*, The Journal of Physical Chemistry B **121**, 683–695 (2017).

[171] M. D. Battistel, R. Pendrill, G. Widmalm, and D. I. Freedberg, *Direct Evidence for Hydrogen Bonding in Glycans: A Combined NMR and Molecular Dynamics Study*, The Journal of Physical Chemistry B **117**, 4860–4869 (2013).

[172] J. Kwon, H. L. Reeves, L. Wang, and D. I. Freedberg, *Revealing elusive conformations of sucrose from hydrogen bond J-coupling in H2O: A combined NMR and quantum mechanics study*, Magnetic Resonance in Chemistry **62**, 742–753 (2024).

[173] L. Babcock and R. Pizer, *Dynamics of boron acid complexation reactions. Formation of 1:1 boron acid-ligand complexes*, Inorganic Chemistry **19**, 56–61 (1980).

[174] C. Miyamoto, K. Suzuki, S. Iwatsuki, M. Inamo, H. D. Takagi, and K. Ishihara, *Kinetic Evidence for High Reactivity of 3-Nitrophenylboronic Acid Compared to Its Conjugate Boronate Ion in Reactions with Ethylene and Propylene Glycols*, Inorganic Chemistry **47**, 1417–1419 (2008).

[175] S. A. Valenzuela, J. R. Howard, H. M. Park, S. Darbha, and E. V. Anslyn, *11B NMR Spectroscopy: Structural Analysis of the Acidity and Reactivity of Phenyl Boronic Acid–Diol Condensations*, The Journal of Organic Chemistry **87**, 15071–15076 (2022).

[176] D. Kusuyama, Y. Samukawa, T. Sugaya, S. Iwatsuki, M. Inamo, H. D. Takagi, and K. Ishihara, *Detailed Reaction Mechanisms of 4-Pyridylboronic Acid and (N-Methyl)-4-Pyridinium Boronic Acid with D-Sorbitol in Aqueous Solution*, ChemistrySelect **4**, 4944–4951 (2019).

[177] G. Springsteen and B. Wang, *A detailed examination of boronic acid–diol complexation*, Tetrahedron **58**, 5291–5300 (2002).

[178] M. G. Evans and M. Polanyi, *Some applications of the transition state method to the calculation of reaction velocities, especially in solution*, Transactions of the Faraday Society **31**, 875 (1935).

[179] L. Bonnet and J. Rayez, *Dynamical derivation of Eyring equation and the second-order kinetic law*, International Journal of Quantum Chemistry **110**, 2355–2359 (2010).

[180] I. N. Levine, *Physical chemistry*, McGraw-Hill Higher Education, Boston [u.a.], 6. ed. edition (2009). Literaturverz. S. 955 - 958.

[181] K. L. Bhat, S. Hayik, and C. W. Bock, *A computational study of the formation of a boron–oxygen–carbon linkage. The reaction of monohydroxy borane with methanol*, Journal of Molecular Structure: THEOCHEM **638**, 107–117 (2003).

[182] K. L. Bhat, S. Hayik, J. N. Corvo, D. M. Marycz, and C. W. Bock, *A computational study of the formation of 1,3,2-dioxaborolane from the reaction of dihydroxy borane with 1,2-ethanediol*, Journal of Molecular Structure: THEOCHEM **673**, 145–154 (2004).

[183] W. L. A. Brooks and B. S. Sumerlin, *Synthesis and Applications of Boronic Acid-Containing Polymers: From Materials to Medicine*, Chem. Rev. **116**, 1375–1397 (2016).

[184] P. Pracht and S. Grimme, *Calculation of absolute molecular entropies and heat capacities made simple*, Chem. Sci. **12**, 6551–6568 (2021).

[185] B. L. Greenstein, D. C. Elsey, and G. R. Hutchison, *Determining best practices for using genetic algorithms in molecular discovery*, The Journal of Chemical Physics **159** (2023).

[186] M. Congreve, R. Carr, C. Murray, and H. Jhoti, *A 'Rule of Three' for fragment-based lead discovery?*, Drug Discovery Today **8**, 876–877 (2003).

[187] A. J. Floyd, S. F. Dyke, and S. E. Ward, *The synthesis of phenanthrenes*, Chemical Reviews **76**, 509–562 (1976).

[188] Y. H. Kim, H. Lee, Y. J. Kim, B. T. Kim, and J.-N. Heo, *Direct One-Pot Synthesis of Phenanthrenes via Suzuki-Miyaura Coupling-Aldol Condensation Cascade Reaction*, The Journal of Organic Chemistry **73**, 495–501 (2007).

[189] J. Liu, S. Tang, M. Zhao, J. Huai, J. Yu, J. Zhao, and P. Li, *Reactivity of Vinyl Epoxides-Oxetanes-Cyclopropanes toward Arynes: Access to Functionalized Phenanthrenes*, ACS Omega **6**, 35852–35865 (2021).

[190] R. Martin and S. L. Buchwald, *Palladium-Catalyzed Suzuki-Miyaura Cross-Coupling Reactions Employing Dialkylbiaryl Phosphine Ligands*, Accounts of Chemical Research **41**, 1461–1473 (2008).

[191] W. C. Fu, Z. Wang, W. T. K. Chan, Z. Lin, and F. Y. Kwong, *Regioselective Synthesis of Polycyclic and Heptagon-embedded Aromatic Compounds through a Versatile π-Extension of Aryl Halides*, Angewandte Chemie International Edition **56**, 7166–7170 (2017).

[192] Y. Zhong, W.-Y. Wu, S.-P. Yu, T.-Y. Fan, H.-T. Yu, N.-G. Li, Z.-H. Shi, Y.-P. Tang, and J.-A. Duan, *A novel and efficient synthesis of phenanthrene derivatives via palladium/norbornadiene-catalyzed domino one-pot reaction*, Beilstein Journal of Organic Chemistry **15**, 291–298 (2019).

[193] T. Hao, W. Gao, S. Yuan, Y. Liu, Y. Li, Y. Fu, and Q. Ding, *One-Pot Three-Component Synthesis of Phenanthrenes via Palladium-Catalyzed Catellani and Retro-Diels-Alder Reactions*, The Journal of Organic Chemistry **88**, 10426–10433 (2023).

[194] B. P. Moloto, P. Vermeeren, M. Dalla Tiezza, C. Esterhuysen, F. M. Bickelhaupt, and T. A. Hamlin, *Palladium-Catalyzed Activation of Carbon-Halogen Bonds: Electrostatics-Controlled Reactivity*, European Journal of Organic Chemistry **2022** (2022).

[195] O. Y. Yuen, S. S. Ng, W. H. Pang, and C. M. So, *Palladium-catalyzed chemoselective Suzuki-Miyaura cross-coupling reaction of poly(pseudo)halogenated arenes*, Journal of Organometallic Chemistry **1005**, 122983 (2024).

[196] F. Beaumard, P. Dauban, and R. H. Dodd, *One-Pot Double Suzuki-Miyaura Couplings: Rapid Access to Nonsymmetrical Tri(hetero)aryl Derivatives*, Organic Letters **11**, 1801–1804 (2009).

[197] S. D. Walker, T. E. Barder, J. R. Martinelli, and S. L. Buchwald, *A Rationally Designed Universal Catalyst for Suzuki-Miyaura Coupling Processes*, Angewandte Chemie International Edition **43**, 1871–1876 (2004).

[198] K. Osakada and Y. Nishihara, *Transmetalation of boronic acids and their derivatives: mechanistic elucidation and relevance to catalysis*, Dalton Transactions **51**, 777–796 (2022).

[199] T. E. Barder, S. D. Walker, J. R. Martinelli, and S. L. Buchwald, *Catalysts for Suzuki-Miyaura Coupling Processes: Scope and Studies of the Effect of Ligand Structure*, Journal of the American Chemical Society **127**, 4685–4696 (2005).

[200] J. Full, S. P. Panchal, J. Götz, A. Krause, and A. Nowak-Król, *Modular Synthesis of Organoboron Helically Chiral Compounds: Cutouts from Extended Helices*, Angewandte Chemie International Edition **60**, 4350–4357 (2021).

[201] A. Nowak-Król, F. Full, M. J. Wildervanck, and D. Volland, *Synthesis of Enantioenriched Azaborole Helicenes by Chirality Transfer from Axially Chiral Biaryls*, Synlett **34**, 477–482 (2022).

[202] F. Full, A. Artigas, K. Wiegand, D. Volland, K. Szkodzińska, Y. Coquerel, and A. Nowak-Król, *Controllable 1,4-Palladium Aryl to Aryl Migration in Fused Systems-Application to the Synthesis of Azaborole Multihelicenes*, Journal of the American Chemical Society **146**, 29245–29254 (2024).

[203] M. J. Jimenez, A. Jaramillo-Botero, and A. Avila, *Au-NP-based colorimetric assay for sugar detection and quantification*, Sensors and Actuators Reports **6**, 100171 (2023).

[204] S. A. Perdomo, V. Ortega, A. Jaramillo-Botero, N. Mancilla, J. H. Mosquera-DeLaCruz, D. P. Valencia, M. Quimbaya, J. D. Contreras, G. E. Velez, O. A. Loaiza,

A. Gomez, and J. de la Roche, *SenSARS: A Low-Cost Portable Electrochemical System for Ultra-Sensitive, Near Real-Time, Diagnostics of SARS-CoV-2 Infections*, IEEE Transactions on Instrumentation and Measurement **70**, 1–10 (2021).

[205] W. J. Mortier, S. K. Ghosh, and S. Shankar, *Electronegativity-equalization method for the calculation of atomic charges in molecules*, Journal of the American Chemical Society **108**, 4315–4320 (1986).

[206] L. Gooßen, N. Rodríguez, P. Lange, and C. Linder, *Decarboxylative Cross-Coupling of Aryl Tosylates with Aromatic Carboxylate Salts*, Angewandte Chemie International Edition **49**, 1111–1114 (2010).

[207] M. Orlandi, D. Brenna, R. Harms, S. Jost, and M. Benaglia, *Recent Developments in the Reduction of Aromatic and Aliphatic Nitro Compounds to Amines*, Organic Process Research and Development **22**, 430–445 (2016).

[208] R. Zangi, *Breakdown of Langmuir Adsorption Isotherm in Small Closed Systems*, Langmuir **40** (2024).

[209] M. Bielecki, H. Eggert, and J. Chr. Norrild, *A fluorescent glucose sensor binding covalently to all five hydroxy groups of α-D-glucofuranose. A reinvestigation*, J. Chem. Soc., Perkin Trans. 2 , 449–456 (1999).

[210] A. Adamczyk-Woźniak, J. T. Gozdalik, D. Wieczorek, I. D. Madura, E. Kaczorowska, E. Brzezińska, A. Sporzyński, and J. Lipok, *Synthesis, Properties and Antimicrobial Activity of 5-Trifluoromethyl-2-formylphenylboronic Acid*, Molecules **25**, 799 (2020).

[211] X. Li, J. Pennington, J. F. Stobaugh, and C. Schöneich, *Synthesis of sulfonamide- and sulfonyl-phenylboronic acid-modified silica phases for boronate affinity chromatography at physiological pH*, Analytical Biochemistry **372**, 227–236 (2008).

[212] F. Li, X. Zhao, and G. Xu, *Synthesis and Application of a Nitrobenzeneboronic Acid-substituated Silica for Affinity Chromatography*, Chinese Journal of Analytical Chemistry **34**, 1366–1370 (2006).

[213] D. Li, Y. Chen, and Z. Liu, *Boronate affinity materials for separation and molecular recognition: structure, properties and applications*, Chem. Soc. Rev. **44**, 8097–8123 (2015).

[214] M. Sanjoh, D. Iizuka, A. Matsumoto, and Y. Miyahara, *Boronate Based Metal-Free Platform for Diphosphate-Specific Molecular Recognitions*, Organic Letters **17**, 588–591 (2015).

[215] A. Matsumoto, A. J. Stephenson-Brown, T. Khan, T. Miyazawa, H. Cabral, K. Kataoka, and Y. Miyahara, *Heterocyclic boronic acids display sialic acid selective binding in a hypoxic tumor relevant acidic environment*, Chemical Science **8**, 6165–6170 (2017).

[216] A. Matsumoto, K. Yamamoto, R. Yoshida, K. Kataoka, T. Aoyagi, and Y. Miyahara, *A totally synthetic glucose responsive gel operating in physiological aqueous conditions*, Chemical Communications **46**, 2203 (2010).