

JÉRÔME RIHON

MOLECULAR MODELING TOOLS TO IMPROVE AND EXPAND
COMPUTATIONAL RESEARCH ON SYNTHETIC NUCLEIC ACIDS

SEPTEMBER 2024

KU LEUVEN

prof. dr. E. Lescrinier
prof. dr. V. B. Pinheiro
prof. dr. M. Froeyen

JÉRÔME RIHON

Dissertation presented in partial
fulfilment of the requirements
for the degree of Doctor in
Pharmaceutical Sciences

KU Leuven
Group Biomedical Sciences
Faculty of Pharmaceutical Sciences
Department Pharmaceutical and Pharmacological Sciences
Laboratory of Medicinal Chemistry
Rega Institute for Medical Research

KU LEUVEN

DOCTORAL SCHOOL
BIOMEDICAL SCIENCES

MOLECULAR MODELING TOOLS TO IMPROVE AND EXPAND COMPUTATIONAL RESEARCH ON SYNTHETIC NUCLEIC ACIDS

Jérôme RIHON

Promotor	Eveline Lescrinier
Co-Promotor	Vitor Bernardes Pinheiro
Co-Promotor	Matheus Froeyen
Chair	Jef Rozenski
Jury member	Kalyan Das
Jury member	Joost Schymkowitz

Dissertation presented in partial
fulfilment of the requirements
for the degree of Doctor in
Pharmaceutical Sciences

September 2024

KU Leuven
Groep Biomedische Wetenschappen
Faculteit Farmaceutische Wetenschappen
Departement Farmaceutische en Farmacologische Wetenschappen
Laboratorium van Medicinale Chemie
Rega Instituut voor Medisch Onderzoek

KU LEUVEN

DOCTORAL SCHOOL
BIOMEDICAL SCIENCES

MOLECULAIRE MODELING TOOLS TER BEVORDERING EN UITBREIDING VAN COMPUTATIONEEL ONDERZOEK OP SYNTHETISCHE NUCLEÏNEZUREN

Jérôme RIHON

Promotor	Eveline Lescrinier
Co-Promotor	Vitor Bernardes Pinheiro
Co-Promotor	Matheus Froeyen
Chair	Jef Rozenski
Jury member	Kalyan Das
Jury member	Joost Schymkowitz

Proefschrift voorgedragen
tot het behalen van de
graad van Doctor in de
Farmaceutische Wetenschappen

ACKNOWLEDGEMENTS

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

ABSTRACT

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

SAMENVATTING

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distingue possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

GLOSSARY

 Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

CONTENTS

ACKNOWLEDGEMENTS	V
ABSTRACT	VI
SAMENVATTING	VII
GLOSSARY	VIII
1. General introduction	1
1.1. Introduction	1
1.1.1. Welcome to the thesis	1
1.2. Methods	1
1.3. Result & Discussion	1
1.4. Conclusion	1
2. The <i>puckers</i> toolkit to calculation puckers values of biomolecular monomers.	2
2.1. Introduction	2
2.1.1. Welcome to the thesis	2
2.1.2. Let's talk about something relevant again	2
2.2. Methods	3
2.3. Result & Discussion	4
2.4. Conclusion	4
2.5. References	6
3. The Ducque model builder tool to predict synthetic nucleic acid duplexes	7
3.1. Introduction	7
3.2. Methods	9
3.2.1. Characterisation of the nucleosides through Quantum Mechanics	9
3.2.2. The Ducque model builder for (X)NA structures	10
3.2.3. Extending the XNA library of Ducque	10
3.2.4. Force field parametrisation for molecular mechanics	11
3.2.5. Molecular Mechanics simulations	12
3.3. Results	12
3.3.1. Charge derivation using ORCA	12
3.3.2. RNA and DNA model building by Ducque	13
3.3.3. MD simulations on Ducque's models validated with (X)NA structures in literature	14
3.3.4. Using the proposed workflow to gain insight in RNA::MNA	14
3.4. Discussion	17
3.5. Conclusion	19
3.5.1. Supplementary information	19
3.6. References	21
4. Curvature Torsion on Synthetic Nucleic Acids	26
4.1. Introduction	26

4.1.1. Welcome to the thesis	26
4.2. Methods	26
4.3. Result & Discussion	26
4.4. Conclusion	26
5. Structure prediction of an HNA aptamer with a G-quadruplex motif	27
5.1. Introduction	27
5.1.1. Welcome to the thesis	27
5.2. Methods	27
5.3. Result & Discussion	27
5.4. Conclusion	27
6. Concluding remarks and future perspective	28
6.1. Concluding remarks	28
6.2. Future perspectives	28
Published works	29
List of Publications	29
Software releases	29
Scientific outreach	29
Declarations	30
Scientific acknowledgements	30
Personal contributions	30
Conflict of Interest	30

GENERAL INTRODUCTION

1

1.1. Introduction

1.1.1. Welcome to the thesis

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

» An addition to the text

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

1.2. Methods

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

1.3. Result & Discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

1.4. Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

THE PUCKERS TOOLKIT TO CALCULATION PUCKERS VALUES OF BIOMOLECULAR MONOMERS.

2

Adapted from the following manuscript *xxx/aaa/xxxxx* :

Rihon J., Reynders S., Pinheiro V.B., Lescrinier E. "The puckers toolkit to facilitate sampling the conformational space of biomolecular monomers." Journal of Cheminformatics (J Cheminf).

2.1. Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

2.1.1. Welcome to the thesis

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distingue possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

› Intermezzo to add something

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut. [1, 2]

2.1.2. Let's talk about something relevant again

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distingue possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedit, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusrusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torquem illum hosti detraxisse, ut aliquam ex eo est consecutus? - Laudem et caritatem, quae sunt

vitae sine metu degendae praesidia firmissima. – Filium morte multavit. – Si sine causa, nolle me ab eo delectari, quod ista Platonis, Aristoteli, Theophrasti orationis ornamenta neglexerit. Nam illud quidem physici, credere aliquid esse minimum, quod profecto numquam putavisset, si a Polyaeno, familiari suo, geometrica discere maluisset quam illum etiam ipsum dedocere. Sol Democrito magnus videtur, quippe homini eruditio in geometriaque perfecto, huic pedalis fortasse; tantum enim esse omnino in nostris poetis aut inertissimae segnitiae est aut fastidii delicatissimi. Mihi quidem videtur, inermis ac nudus est. Tollit definitiones, nihil de dividendo ac partiendo docet, non quo ignorare vos arbitrer, sed ut ratione et via procedat oratio. Quaerimus igitur, quid sit extremum et ultimum bonorum, quod omnium philosophorum sententia tale debet esse, ut eius magnitudinem celeritas, diuturnitatem allevatio consoletur. Ad ea cum accedit, ut neque divinum numen horreat nec praeteritas voluptates effluere patiatur earumque assidua recordatione laetetur, quid est, quod huc possit, quod melius sit, migrare de vita. His rebus instructus semper est in voluptate esse aut in armatum hostem impetum fecisse.

2.2. Methods

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequa doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distingue possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos iridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitibus aut rerum necessitatibus saepe eveniet, ut et voluptates re-pudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusrusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torque illum hosti detraxisse, ut aliquam ex eo est consecutus? – Laudem et caritatem, quae sunt vitae sine metu degendae praesidia firmissima. – Filium morte multavit. – Si sine causa, nolle me ab eo delectari, quod ista Platonis, Aristoteli, Theophrasti orationis ornamenta neglexerit. Nam illud quidem physici, credere aliquid esse minimum, quod profecto numquam putavisset, si a Polyaeno, familiari suo, geometrica discere maluisset quam illum etiam ipsum dedocere. Sol Democrito magnus videtur, quippe homini eruditio in geometriaque perfecto, huic pedalis fortasse; tantum enim esse omnino in nostris poetis aut inertissimae segnitiae est aut fastidii delicatissimi. Mihi quidem videtur, inermis ac nudus est. Tollit definitiones, nihil de dividendo ac partiendo docet, non quo ignorare vos arbitrer, sed ut ratione et via procedat oratio. Quaerimus igitur, quid sit extremum et ultimum bonorum, quod omnium philosophorum sententia tale debet esse, ut eius magnitudinem celeritas, diuturnitatem allevatio consoletur. Ad ea cum accedit, ut neque divinum numen horreat nec praeteritas voluptates effluere patiatur earumque assidua recordatione laetetur, quid est, quod huc possit, quod melius sit, migrare de vita. His rebus instructus semper est in voluptate esse aut in armatum hostem impetum fecisse aut in poetis evolvendis, ut ego et Triarius te hortatore facimus, consumeret, in quibus hoc primum est in quo admirer, cur in gravissimis rebus non delectet eos sermo patrius, cum idem fabellas Latinas ad verbum e Graecis expressas non invitit legant. Quis

enim tam inimicus paene nomini Romano est, qui Ennii Medeam aut Antiopam Pacuvii spernat aut reiciat, quod se isdem Euripidis fabulis delectari dicat, Latinas litteras oderit? Synephebos ego, inquit, potius Caecilii aut Andriam Terentii quam utramque Menandri legam? A quibus tantum dissentio, ut, cum Sophocles vel optime scripserit Electram, tamen male conversam Atilii mihi legendam putem, de quo Lucilius: 'ferreum scriptorem', verum, opinor, scriptorem tamen, ut legendus sit. Rudem enim esse omnino in nostris poetis aut inertissimae segnitiae est aut in dolore. Omnis autem privatione doloris putat Epicurus.

2.3. Result & Discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequa doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distingue possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitibus aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedit, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusrusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torque illum hosti detraxisse, ut aliquam ex eo est consecutus? – Laudem et caritatem, quae sunt vitae sine metu degendae praesidia firmissima. – Filium morte multavit. – Si sine causa, nolle me ab eo delectari, quod ista Platonis, Aristoteli, Theophrasti orationis ornamenta neglexerit. Nam illud quidem physici, credere aliquid esse minimum, quod profecto numquam putavisset, si a Polyaeno, familiari suo, geometrica discere maluisset quam illum etiam ipsum dedocere. Sol Democrito magnus videtur, quippe homini erudito in geometriaque perfecto, huic pedalis fortasse; tantum enim esse omnino in nostris poetis aut inertissimae segnitiae est aut fastidii delicatissimi. Mihi quidem videtur, inermis ac nudus est. Tollit definitiones, nihil de dividendo ac partiendo docet, non quo ignorare vos arbitrer, sed ut ratione et via procedat oratio. Quaerimus igitur, quid sit extreum et ultimum bonorum, quod omnium philosophorum sententia tale debet esse, ut eius magnitudinem celeritas, diuturnitatem allevatio consoletur. Ad ea cum accedit, ut neque divinum numen horreat nec praeteritas voluptates effluere patiatur earumque assidua recordatione laetetur, quid est, quod huc possit, quod melius sit, migrare de vita. His rebus instructus semper est in voluptate esse aut in armatum hostem impetum fecisse aut in poetis evolvendis, ut ego et Triarius te hortatore facimus, consumeret, in quibus hoc primum est in quo admirer, cur in gravissimis rebus non delectet eos sermo patrius, cum. [3]

2.4. Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequa doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distingue possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Sto-

icos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitibus aut rerum necessitatibus saepe eveniet, ut et voluptates re-pudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non de-pravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis cho-rusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torquem illum hosti detraxisse, ut aliquam ex eo est consecutus? – Laudem et caritatem, quae sunt vitae sine metu degendae praesidia firmissima. – Filium morte multavit. – Si sine causa, nolle me ab eo delectari, quod ista Platonis, Aristoteli, Theophrasti orationis ornamenta neglexerit. Nam illud quidem physici, credere aliquid esse minimum, quod profecto numquam putavisset, si a Polyaeno, familiari suo, geometrica discere maluisset quam illum etiam ipsum dedocere. Sol Democrito mag-nus videtur, quippe homini erudito in geometriaque perfecto, huic pedalis fortasse; tantum enim esse omnino in nostris poetis aut inertissimae segnitiae est aut fastidii delicatissimi. Mihi quidem videtur, inermis ac nudus est. Tollit definitiones, nihil de dividendo ac partiendo docet, non quo ig-norare vos arbitrer, sed ut.

2.5. References

- [1] Cremer. D., Pople. J., "General definition of ring puckering coordinates" *Journal of the American Chemical Society*, vol. 97, pp. 1354–1358, Mar. 1975, NO DOI
- [2] IUPAC-IUB. J., "Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains" *Journal of Pure and Applied Chemistry*, vol. 55, pp. 1273-1280, Aug. 1983, NO DOI
- [3] Neese. F., Wennmohs. F., Becker. U., Ripplinger. C., "The ORCA quantum chemistry program package" *The Journal of Chemical Physics*, vol. 152, pp. 224108, Jun. 2020, NO DOI

THE DUCQUE MODEL BUILDER TOOL TO PREDICT SYNTHETIC NUCLEIC ACID DUPLEXES

3

Adapted from the following manuscript 10.1093/nar/gkae135 :

Rihon J., Mattelaer C.-A., Montalvão R.W., Froeyen M., Pinheiro V.B., Lescrinier E. “*Structural insights into the Morpholino Nucleic Acid/RNA duplex using the new XNA builder Ducque in a molecular modeling pipeline.*” Nucleic Acid Research (NAR).

3.1. Introduction

The potential of nucleic acid therapeutics was recognized more than 50 years ago, but nucleases that hydrolyse phosphodiester linkages in oligonucleotide (ON) strands represented one of the hurdles that significantly slowed down their rise in the therapeutic field [1]. Analogue carrying modifications in the backbone and nucleobase (xenobiotic nucleic acids, XNAs [2]) accelerated development of nucleic acids therapeutics in the mid-1980s, as they allowed to modulate catalytic degradation of ONs and/or increase affinity to the selected target. Several XNA chemistries were used for various clinical applications, of which some have reached the market [3]. While XNA were initially designed to modulate RNA activity upon hybridisation, some XNA alterations have also been developed for applications such as therapeutic aptamers and an alternative genetic system in synthetic biology. This has lead to the engineering of enzymes able to process such ON strands, in order to promote XNA ON manipulation methodologies to a higher level of efficiency, for both researcher and cell [4, 5].

Conformational preorganization of sugar moieties in the XNA backbone is a key feature for their binding capacity to native nucleotides. Chemical modifications such as pyranose units [6] and locked nucleic acids (LNA) [7] freeze the sugar moiety in the nucleic acid backbone and offer the ultimate control of conformational preorganization and fix their recognition potential for either DNA or RNA. The xylose-based chemistries, on the other hand, avoid hybridisation with native nucleotides [8]. Nucleobase alterations can provide hybridisation stability, which benefits duplex formation, or extends binding potential, broadening the range of possible aptamer-target interactions. Recently, these XNA have been successfully employed in CRISPR-Cas9 research, using the thG modification [9].

The morpholino nucleic acids (MNAs) are a prime example of highly modified nucleic acids in antisense oligonucleotide (ASO) research and therapies [10, 11]. Their presumed mechanism of action is to sterically block RNA splicing or translation, by binding to their RNA complement. Preclinical studies in zebrafish have popularised the use of such gene-knockdown mechanisms with MNA ONs [12]. MNA are commonly used in the treatment for Duchenne’s Muscular Dystrophy, where binding to the specific exon-51 sequence leads to trimming of the dystrophin protein. In short, the binding of an ASO to its target sequence results in cancelling the propagation of that exon-51 RNA sequence into a peptide strand, thereby halting the translation into a full-fledged protein. This is in contrast with the endogenous RNA degradation exploited by other XNA chemistries for ASO and silencing RNA (siRNA) purposes [13, 14, 15] catalysed by the RNase H enzyme or RISC complex, that break RNA strands upon hybridisation with complementary DNA and RNA strands respectively. In such a mechanism, the XNA::RNA duplex must be recognised by the processing machinery through simi-

larity to the natural substrate [16]. Experimental structures of a range of XNA chemistries in homoduplexes and bound to a complementary nucleic acid sequence proved useful in understanding their biophysical and biological properties [5].

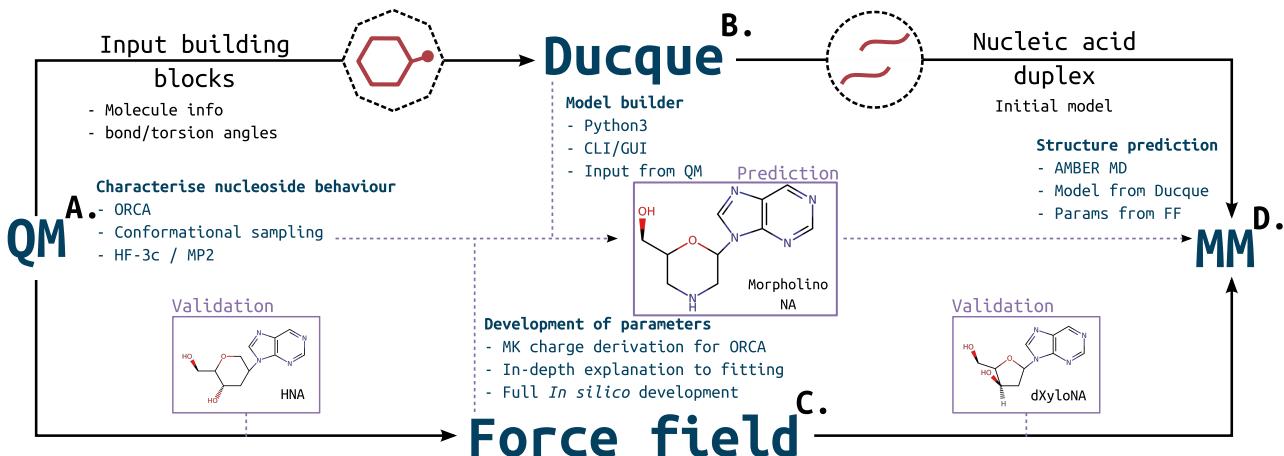


Figure 3-1: **A.** QM approach to perform conformational sampling to describe the torsional behaviour. **B.** The Ducque model builder receives input from the computed PES to use curated conformers as building block to generate duplex models. **C.** From the PES, one can derive a force field by virtue of the predicted behaviour. A charge derivation scheme has been implemented for ORCA. Curated conformers are to be used to derive torsional parameters for the force field. **D.** Combining the products of (B., an (X)NA molecular model) and (C., a representative force field), we can predict the molecular structure of an XNA model duplex through the use of an MM package, which can be used for antisense design, (X)NA enzyme engineering and more. This project workflow has been validated on HNA and dXyNA and has been applied to predict the RNA::MNA heteroduplex.

Molecular modeling of XNA::RNA duplexes can predict possible XNA applications and guide the design of XNA and XNA processing enzymes [17]. Model building nucleic acids has been facilitated in the mid '90s thanks to the release of the Nucleic Acid Builder (NAB) software and its domain-specific language [18]. This was followed with many implementations of the NAB language in a variety of wrappers and servers to allow model building for DNA and RNA. The proto-Nucleic Acid Builder has taken a small step towards XNA building by including a set of modifications known to build into mainly A-type structures [19]. It uses parameters from experimentally derived structures, using the convention of 3DNA [20] of basis reference frames, which differ slightly from the definition of Curves+ [21]. It also makes a good attempt at minimising duplex models with force fields (FFs) that were developed for small molecule minimisation and evaluation. The claimed predictions simply build along a set of given vectors from data gathered from wet lab experiments. Consequently, such building schemes are not applicable for many unsolved and underresearched chemistries structures.

Because of the amount of new chemistries developed for many different applications [22, 23], an urgent need arised to design a platform that can handle the influx of existing XNAs and modifications that have yet to be developed. With this in mind, we developed the Ducque software from scratch, independent from the approach of the traditional Nucleic Acid Builder. It comes with a user costumisable library for linker and sugar moieties, reserving functionality for purine- and pyrimidine-like bases, to build XNA models with a virtually unlimited set of chemistries on the backbone, with a future outlook on customising nucleobase modifications. This approach easily allows to build all models exemplified in Anosova *et al.* [24], regardless of complexity.

The Ducque model builder is part of an elaborate workflow to implement new chemistries for accurate molecular dynamics (MD) simulations (Figure 3-1). The model builder also boasts a neat graphic user interface (GUI) that makes the flow of importing new building blocks, randomising and build-

ing structures smooth and concise, by allowing access to the various modules in a chained fashion. %The Ducque model builder is part of an elaborate workflow to implement new chemistries for accurate molecular dynamics (MD) simulations (Figure WORKFLOW). The model builder also boasts a neat graphic user interface (GUI) that makes the flow of importing new building blocks, randomising and building structures smooth and concise, by allowing access to the various modules in a chained fashion. A potential energy surface (PES) describes the nucleoside's behavior along its puckering profile. This data feeds directly into Ducque and the creation of a suitable FF, respectively. Additionally, an AMBER-specific implementation of the parametrisation scheme was developed for the ORCA package. While Ducque is used to build an initial model, the force field is essential for the final Molecular Mechanics (MM) stage that provides an accurate structure. Here, we demonstrate Ducque against natural nucleic acids using NAB as our benchmark and show its functionality across a range of well-characterised XNAs. The workflow was validated on Hexitol Nucleic Acid (HNA) and deoxy-xylose nucleic acid (dXyNA), and challenged on the RNA::MNA duplex, a heteroduplex with clinical relevance for which no experimental structure is available to date.

3.2. Methods

3.2.1. Characterisation of the nucleosides through Quantum Mechanics

The conformational sampling method has been previously reported and applied on DNA, RNA, HNA and dXyNA, in Mattelaer *et al.* [25]. For MNA, the Cremer-Pople (CP) coordinates can be reversed engineered to Strauss-Pickett (SP) improper dihedrals sugarpucker, by generating equidistributed coordinate points on the surface of a sphere [26]. The generated spherical coordinates (r, θ, φ) are converted to local elevation (z_i) from a mean plane per respective atom [27], by assuming bond lengths and angles, which positions the atoms in the ring in cartesian space. From the atomic coordinates, the improper dihedrals ($\alpha_1, \alpha_2, \alpha_3$) can be calculated for [28]. Graphical representations of these formalisms have been added (Figure CPMeanplane, SPIImproper). These improper dihedrals are used as restraints for the ORCA software (v. 5.0.2) [29]. Several backbone torsion angles were restrained as follows : β (180.1°), γ (60.1°), ε (180.1°), χ (193.9°) [25]. First, a lightweight semiempirical HF-3c (Hartree Fock) geometry optimisation is performed to relax the respective conformations after imposing the restraints. Afterwards, a single point energy (SPE) evaluation, through Møller-Plesset 2nd order perturbation theory (MP2) is performed to evaluate the potential of the respective conformers. The specific accompanying basis set for MP2 is 6-311++G(2df,2p) with the def2-qzvpp/ C basis as an auxiliary basis set for the RI approximation and the def2/JK auxiliary basis set for the RI approximation of Coulomb and exchange integrals in the HF step [25]. A total of 630 conformers were sampled for each morpholino nucleoside. The geography library, *cartopy* (\leq python3.6), is employed to make a 2D projection of the \mathbb{R}^3 surface of the conformational sphere of HNA and MNA from which we sampled (Figure spherical_convention). The Mollweide projection is used, together with a PlateCarree transformation. The SciPy Radial basis function (Rbf) is used to interpolate the values to generate the PES. QM calculations were performed on a Ryzen Threadripper 3970, with a memory capacity of 32GB RAM. The standard MNA has been altered to a methyl-capped O6' and N3' variant before sampling the conformations, for reasons detailed in the **Results and Discussion**.

The linker moiety, N,N-dimethyl-O-methylaminophosphoroamide, has been subjected to a conformational sampling following the same methodology, but instead of varying endocyclic torsion

angles, the (ζ, α) dihedral dyad was varied from 0° to 360° in steps of 10° along both axes (Figure PESLinker), for a total of 1369 conformers.

3.2.2. The Ducque model builder for (X)NA structures

Ducque is written in Python3 and uses only built-in modules as well as NumPy-based modules (NumPy, SciPy) [30]. Ducque is functional both in the Commandline Interface (CLI) and with a GUI (Tkinter). It is free and open-source software (jrihon/Ducque), under the MIT license.\ Ducque uses a mechanical building method that functions solely on the given nucleoside conformers and backbone parameters of the respective nucleotides it builds into a strand, meaning it only takes into account the input bond angles and dihedrals. Array rotation in \mathbb{R}^3 is supported by quaternion mathematics.

Ducque's core functionality is extrapolation of the backbone dihedral by requiring only torsion and angle values. To position an atom in \mathbb{R}^3 , there is only one location that satisfies both the given angle and dihedral. With this process, and an associated fitting of the following linker and nucleoside, the leading strand is built (Figure DihedralExtrapol). To build to complementary sequence, a fitting of the complementary strand's nucleobase to the leading strand's nucleobase is performed, by posing the complementary nucleobase onto the plane of its basepair and shifting it into the desired hydrogen bonding position (Figure BuildComp). An added fitting of the backbone with neighbouring residues resolves most, if not all, clashes. Any remaining clashes are easily resolved through minimisation in the next stage of MD, with the correct force field.

3.2.3. Extending the XNA library of Ducque

Ducque allows the user to implement a custom repository of new chemistries, which function as building blocks for the virtual duplexes. This module requires the additional information on the input angles and the *pdb* format nucleoside itself. Furthermore, one can generate randomised sequences through the randomisation module, or manually customise sequences, giving the user the freedom of choice when designing a duplex, e.g. multiple chemistries. Lastly, the *xyz* format files, generated during Conformational Sampling by ORCA, can be converted to a *pdb* format by Ducque. This implies batch conversion through Ducque as well.

For the extension of the Ducque library with new inputs, a *pdb* inputfile of the new nucleoside must be provided by the user. It is converted to the *json* format (--transmute), which requires parameters like residue name, backbone angles and dihedrals to generate a suitable textit{json"}, biblio) inputfile. The *xyz* format files, generated during conformational sampling to characterise nucleosides (ORCA output), can be converted to *pdb* format files by Ducque (--xyz_pdb) and subsequently used to supply to the user library. One can generate strands with randomised sequences for a specific XNA chemistry through the randomisation module (--randomise), or manually customise sequences. The latter option allows to design a nucleobase sequence as well as the nucleic acid backbone on a residue specific level, giving freedom of choice for duplex design with multiple chemistries in the (customised) library. The randomisation functionality was used to generate a multitude of different models for DNA::DNA, DNA::RNA, RNA::RNA, dXyNA::dXyNA, RNA::HNA and finally RNA::MNA. A sequence was specified manually to generate double-Drew Dickerson dodecamer (DDD) structures. Nucleic acid duplex models were built (--build) from the randomised/DDD sequences and were subjected to MD.

The respective nucleobases were flattened using the HF-3c level geometry optimisation, with the atoms in and on the sugar restrained in cartesian space and the math.chi - χ dihedral restrained at its current position. The improper dihedral for purines (C1', C8, N9, C4) and pyrimidines (C1', C6, N1, C2) is also restrained to -179.5° math.degree to ensure a plane as flat as the standard nucleobases provided with NAB to satisfy the planarity assumptions in the rotations. For benchmarking the build quality of the structures, python scripting (for NAB) and Ducque's randomisation module were used to generate ten RNA and DNA homoduplexes of a randomised sequence of 12 basepairs (bp), respectively. These were compared to duplexes generated by the NAB language (Figure BenchmarkNA), Table BenchmarkNA). The `time` shell command (logging the *real* output) was used to compare build times for chronometric benchmarking. NAB was logged for combined time of compiling from the Domain-specific Language (DSL) to the binary and executing the binary, while Ducque can only be logged for Python's runtime. The structures were evaluated based on RMSD and difference in inter-basepair-parameters using Curves+ [31].

3.2.4. Force field parametrisation for molecular mechanics

%Two parameters define a force field : a set of charges and the geometrical properties of the molecules. The AMBER FF [32, 33] is compatible with restraint electrostatic potential (RESP) charges [34]. This is a subset of molecular electrostatic potential (MEP) charges sampled with the Merz-Singh-Kollman (MK) population analysis scheme [35, 36], followed by a two-stage restraint procedure to derive atomic charges (RESP fitting). The nucleosides were superposed on the X-Z cartesian plane by the (C1', C5', N3') atoms in the nucleosides through an in-house scripted reorientation [37].

The MK population analysis scheme has been implemented, by the authors, for ORCA and detailed here below. Geometry optimisation of the molecules was done at the PBE0 level, which was followed by an SPE at the HF/6-31G* level/basisset, with ORCA, to generate the correct molecular density orbitals for RESP [38]. The gridpoints are generated through the MSMS software [39] at a range of factors times (1.4, 1.6, 1.8 and 2.0) the van der Waals (vdW) radius, defined by the Connolly surface algorithm [40]. The probe radius was set to 1.4 Å, the density of the points on the surface was set to 3 /Å² (Figure MKPopAnalScheme). MSMS employs a further triangulation of the vertices to distribute them more equally. ORCA's `orca_vpot` program is used to map the orbitals and their densities to the respective gridpoints, in atomic units (a.u.). Afterwards, the ESP-loaded grid is provided to the RESP script for a two-stage restraintive procedure, deriving point charges for the atoms. Similar atoms are equivalenced and internal fragments are restraint to a set charge for the respective nucleosides, which ultimately returns the needed RESP charges for the FF [34, 38, 41].

This uses the Chirliam-FrancI least square fitting algorithm [42] (Figure ESPGrid). This methodology was compared against results from GAMESS [43] (Figure MKPopAnalScheme), to compare the grid generation, and against the R.E.D. server (Figure CompareChargedervation), to compare charge derivation. With respect to the nucleosides, all atoms in the ring and all substituents were set for equivalencing, except for C1' and H1' which should be equivalenced with the nucleobases on the occasion of multiple conformers. Hydrogens belonging to amines of nucleobases are also equivalenced, as are the OP1 and OP2 oxygens in the phosphoramidate linker. For second stage RESP, all methyls and methylenes were equivalenced, due to degenerate hydrogens. The force field of MNA contains charges for 6'-head, 6'-fragments-3', 3'-tail, neutral and methyl-capped MNA. Only the lowest energy ⁴C₁ conformer was considered for charge derivation. To create terminal residues and in-

strand fragments, 6' and 3' ends were appropriately restrained with the linker to have a net charge of zero.

Finally, all bond, angle and torsion parameters for the morpholino chemistry were manually curated from the *parm10.dat* frcmod file, contained within the AMBER package, of terms that already described parts of the molecule well, before fitting to the conformers. Only torsion terms of the morpholino ring moiety were fitted for. Paramfit [44, 45] was used to equate QM to MM energies correctly, which also produces a valid *frcmod* file. All conformers in Table relEnergyConformers) were initially considered for fitting torsion angles. Finally, only ${}^4\text{C}_1$, ${}^2\text{T}_3$, ${}^2\text{S}_4$, ${}^4\text{S}_2$, ${}^3\text{T}_2$ of the MNA cytosine were employed for fitting.

A customised CF atom type was created, which was made equivalent to the standard CT atom type, common in AMBER FFs. This was done so as not to override any parameters when loading in the MNA FF together with the standard FFs.

3.2.5. Molecular Mechanics simulations

The Molecular Dynamics simulations were run using the AMBER18 software package [32, 33] joined with AMBERTools19, employing the Particle Mesh Ewald (PMEMD) simulation engine [46]. The MNA force field was imported into LEaP, together with the DNA.OL15 and the RNA.OL3 FF [47, 48]. The TIP3P water model is used for the explicit solvation, in a truncated octahedron box, and the charges were neutralised. dXyNA parameters [8] use the DNA.OL15 FF, HNA parameters were derived from a similar methodology as the one described in this article [49]. Torsion angles for the linker were taken from the GAFF2 parameters, in combination with phosphate parameters from *parm10.dat*. A cutoff distance of 12 Å was used for non-bonded interactions. The minimisation ran for a total of 30 000 cycles, with the first 22 500 cycles employing the steepest descent method and the last 7 500 the conjugate gradient method. The SHAKE algorithm [50] was employed to allow a timestep of 2 fs. An initial heating was performed for 50 ps, from 0 K to 100 K with *vlimit* set to 15. The rest of the heating, from 100 K to 300 K, ran for an additional 50 ps. Density and equilibration ran for 100 ps each, with density set at 1 g/mL. The Langevin thermostat [51] and the Berendsen barostat [52] were used to keep the temperature and density at a constant value. The production simulations were run for 200 ns, unless stated otherwise. Trajectory analysis of the simulations calculated through Cpptraj [53]. MD simulations were run on NVIDIA GeForce RTX 2070 and 2060 through the textit["cuda"], biblio) accelerated simulation engine [54, 55]. For all duplexes, distance-based NMR restraints were employed for the first 15 ns. For structural comparison, superpositioning of the structures was done with UCSF Chimera's MatchMaker tool and all structures were entirely fitted according to their secondary structure. The Curves+ software was used to extract interbasepair data [31].

3.3. Results

3.3.1. Charge derivation using ORCA

The ORCA QM package does not contain the MK population analysis scheme. To keep the workflow free, local and easy to use, we implemented the population analysis scheme for ORCA to generate ESP charges that are ready to use for the RESP fitting. Two programs were used to compare Connolly surface grid generation to decide which program is most suited for the task at hand. The AMBERTools-script *molsurf.c*, which calculates molecular surface areas, was modified to output the vertices of the surface points by which it computes surface areas. A second program, MSMS was also tested (Scripps Research Institute, utilised by VMD [56] and UCSF Chimera [57]). The latter com-

putes a further triangulation of the solvent excluded surfaces (SES) to generate an even distribution of points across the surface of the molecule. Upon visual inspection and level of customisability of the grid (atomic and probe radii, point density), MSMS was hugely favoured. Further comparison of the method was done against GAMESS' output, which defaults to a grid with a density of $1/\text{\AA}^2$. Our implementation for ORCA uses a density of $3/\text{\AA}^2$ [35, 36]. Point density of $1/\text{\AA}^2$ was used for our method to compare against GAMESS. Our methodology shows an even distribution of the surface, compared to GAMESS, and decided to proceed with this approach as it produced a near-identical range of ESP values to that of GAMESS' (Figure MKPopAnalScheme and ESPGrid). A benchmarking against RESP-charge derivation by the R.E.D. server [58] was carried out on HNA nucleosides. We see near-identical atomic charges being fitted to the atoms in the respective nucleosides (Figure CompareChargedervation). Data points from either methodology were compared by pairwise atom alignment (linear regression, SciPy) per nucleoside. This results in an r^2 of 0.985 for hA, 0.985 for hC, 0.987 for hG and 0.979 for hT.

3.3.2. RNA and DNA model building by Ducque

A-type and B-type duplexes of DNA and RNA (Fig. SIShowcase A1.) have been built by Ducque and its performance was compared to NAB to assess its model building capacity. Conserving the planarity of the nucleobases is necessary because Ducque's placement and fitting of the nucleosides in the strands assumes planarity of the nucleobases (Figure BuildComp). Because of their sp^3 type hybridisation, amines tend towards slight tetrahedral structures in nucleobases, and this gets more pronounced in *ab initio* calculations at MP2 levels with midsized basis sets for geometry optimisations. This compromises the flatness of the nucleobase by several tenths of degrees for the nitrogen involved in the glycosidic bond, to several degrees for the amine on the six-membered ring in purines [59]. Optimising at the current level compromises the planarity assumption, introducing distortions during duplex building and causes peculiar rotations of nucleotides in the final model. *Nota bene*, using larger basis sets, i.e. reaching complete basis set, decreases the degree of pyramidalisation [60]. Using this basis would be too computationally intensive for our needs and would only finetune details that would be negated in the following simulations. Therefore, the planarity issue of nucleobases was resolved by the use of planarity restraints on nucleobases during HF-3c geometry optimisations, providing nucleosides for the Ducque library with planar nucleobases, like in NAB.

The quality of the model building is assessed by superpositioning generated models from NAB [18] vs. Ducque (Table BenchmarkNA and Figure BenchmarkNA), which was done across all (twelve) residue pairs on heavy atoms, for ten models of DNA and RNA each per software. For DNA, mean RMSD value is 0.011 \AA (± 0.005) with the lowest value being 0.004 \AA . For RNA, mean RMSD value is 0.169 \AA (± 0.042) with the lowest value being 0.109 \AA . The near-zero RMSD showcases Ducque's qualitative output. For the interbasepair-parameters, we see that the Ducque structures are in good agreement with NAB's. Only the roll parameter, for RNAs, seems to slightly differ. This stems from the fitting protocol not rolling all nucleobases with the exact same value, as this feature is not hard-coded. The simulated structure for the DNA and RNA homoduplex experiments, generated by both NAB or Ducque, show equivalent RMSD trajectories, as is to be expected given the similarity of the starting structures. Since all structures minimised without a problem, both the comparison of structural parameters and the following simulations indicate the quality of the generated structures. With respect to wallclock building-time, NAB averaged 0.272 s and 0.245 s for DNA and RNA respectively.

Ducque, averaging 0.328 s and 0.234 s for DNA and RNA respectively, takes infinitesimally longer for DNA as it considers multiple conformers during the fitting (benchmarked with Python3.12).

3.3.3. MD simulations on Ducque's models validated with (X)NA structures in literature

To assess the quality of Ducque's models for MD simulations, a series of duplexes were studied, starting with (deoxy)-ribose based nucleic acids, which employed the standard AMBER DNA.OL15 and RNA.OL3 FFs. The simulated structures for the DNA and RNA homoduplex experiments, generated by both NAB or Ducque, show equivalent RMSD trajectories (Table BenchmarkNA and Figure BenchmarkNA), as expected given the similarity of the starting structures. An MD simulation on a B-type DNA::RNA and A-type RNA::DNA duplex of identical sequence both built by Ducque, converged into the same A-like structure, as experimentally determined in Davis textit{et al."}, biblio) [61] (Figure ref{fig:RNADNA"}, biblio)). A second experiment was performed with the dXyNA homoduplex. This started out from a typical ladder-like structure [8]. During the MD simulation, the initial structure of the duplex undergoes a remarkable transition to a highly dynamic, left-handed helix (Figure Dxyl0, DistrNormDxyna, Table DxynaBackbone) that matches the type II duplex previously described for xylose based nucleic acids [62, 63, 64], having significantly different backbone dihedral angles compared to those in the initial model. Thirdly, an RNA::HNA duplex was built, using the lowest energy $^4C_1'$ conformations of the PES [25] of the four nucleobases and were prepared and loaded into Ducque's library. A subsequent MD simulation resulted in a stable duplex that matches crystallographic results on sugar puckering as well as overall helical structure [65] (Figure HNA duplex). The methodology of deriving torsional parameters for HNA has been validated already [49], which has also been applied to deriving the parameters for the MNA chemistry.

3.3.4. Using the proposed workflow to gain insight in RNA::MNA

› Quantum Mechanics to characterise the nucleosides and linker

Following the procedure described by Mattelaer *et al.* [25], PESs were generated for the morpholino nucleosides with adenine, guanine, thymine and cytosine nucleobases. We observed intramolecular hydrogen bonds in the initial calculations, inflating equatorial conformations towards higher energy potentials (Figure PESunrestrained). Those observed hydrogen bonds, which would not exist in a polymeric molecule, introduce a forced steric strain on the nucleoside that would not yield representative FF parameters that are applicable for simulations on oligonucleotides. For the capped MNA variant (Figure PESmethylMNA, Table relEnergyconformers), we see a global minimum on the north pole of the CP-sphere, a $^4C_1'$ configuration for all four nucleosides. Local minima are observed mainly on the upper hemisphere, separated by large peaks corresponding to envelope configurations. Equatorial puckering modes tend to be more energetically favourable, with boats and skewwts alike. A final local minimum is observed on the southern hemisphere, in the East.

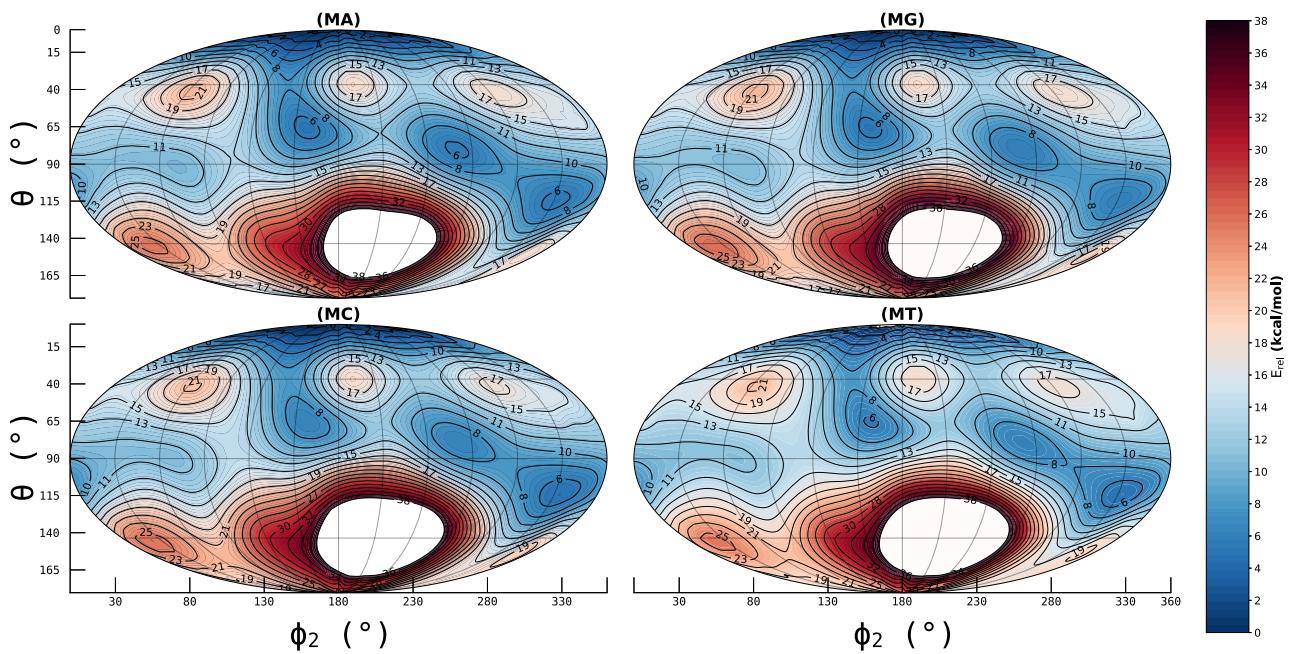


Figure 3-2: Potential Energy Surface of the Morpholino Nucleic Acid. Methyl-caps substituted the HO^{6'} and NH^{3'} on the O^{6'} and N^{3'} atoms respectively, to mimic estimated backbone angles and prevent internal hydrogen bonds from occurring during geometry optimisations.

The N,N-dimethylaminophosphoramidate represents the linkage between successive morpholino nucleosides, where the N,N-dimethylamino mimics the effect of the preceding morpholino ring. A sampling of the (ζ , α) dihedral dyad was carried out too. Minima found at (t, g⁺), (t, g⁻), (g⁻, g⁻), (g⁺, g⁺) and (g⁻, ap⁺), Figure PESlinker. A hard ridge ($\alpha = 180^\circ \pm 30$) splits the PES in two. Its cause is attributed to steric clashes during the sampling itself, due to methyls on either end of the linker residing in close proximity of one another at α in the *anti* range.

› Force field parametrisation

The PESs of the four morpholino nucleosides successfully provided valuable information on transitional pathways. By using the improper dihedrals (α_1 , α_2 , α_3) as restraints on the pyranose ring, we gained information of possible transitions of conformations as well as puckers we should be expecting. Conformations highlighted in Table relEnergyconformers are conformations that were initially considered for parametrisation, specifically for fitting of the sugar torsion angles (see Fig. sphericalconvention for a reference to their respective physical location on the PES).

For the linker fragment, the minima and transition (g⁻, g⁺) conformers of the lower half were used to derive partial charges. Since upper half conformers are mirror images of the lower half, they were not considered. The linker torsional parameters were obtained from GAFF2 [66] and are in good agreement with the PES. (Figure PESlinker C.). The (ζ , α) dihedral dyad values were extracted from the simulation and categorised into discrete integer values, for then to be normalised and plotted. This shows a clear and single peak at the (t, g) range, which is in contrast with the standard phosphate, that stabilises in the (g⁻, g⁺) range in a DNA or RNA duplex [67]. This validates usage of the GAFF2 parameters for the linker. *Nota bene*, a single peak is attributed to the conformational constraints that the linker is submitted to in a polynucleotide chain, which clarifies why multiple minima are perceived in the PES, but only one can be inhabited.

› Molecular model of RNA::MNA

The RNA::MNA duplex was built in Ducque starting with a 24-mer RNA sequence corresponding to a double-Drew-Dickerson-Dodecamer (double-DDD : 5'-CGCGAAUUCGCGCGCGAAUCGCG-3', residues 1-24) with the geometry of an A-type duplex. All ${}^4\text{C}_1$ - MNA nucleosides were ported into Ducque and queried as the complementary strand (residues 25-48). The obtained initial structure was subjected to an MM simulation. An RNA::RNA duplex with identical sequence was generated and simulated for comparative purposes. Results of the simulation are summarised in Figure ResultFINALSIMRNAMNA. In the last 50 ns of the MD trajectory, a pronounced prevalence for the ${}^3\text{E}$ conformation in the conformational wheel is observed for RNA nucleosides in both the dsRNA and heteroduplex model (Figure ResultFINALSIMRNAMNA C,D, Table backbone). The puckering modes for the morpholino rings observed on the CP-sphere (Figure ResultFINALSIMRNAMNA C), in the same timeframe as stated earlier, largely correspond to the low energy regions of the PES (Figure PES_methylMNA", biblio)) The scattered, equatorial conformers correspond to residues 30-31. Their appearance correlate with a K^+ -ion that had been enclosed by two MNA residues (30-31) during the second half of the simulation. Its binding induced local backbone angles to deviate substantially from the other nucleosides, off-setting average and standard deviations (std.) parameters by several degrees. Up to four (25-28) residues had frayed (forming *C*, *E* and *B* conformers alike), causing residue 28 with guanine nucleobase (mG28) to have its β angle to twist into an unfavourable position. Structural changes due to K^+ -binding destabilised downstream stacking, which causes mA30 to lose pairing with its complement. This resulted in a nearby K^+ -ion to alter the χ angle's orientation, of the freed nucleotide, from *anti* to *syn* [68]. The basepairing of mA30 with its complement was reinstated later, in Hoogsteen (HG), with the ion bound (Figure IonBackbone). The K^+ ion is coordinated by mT31:O6', mT31:O5' and mA30:N3' in the MNA backbone and N3 of the nucleobase in mA30. Coordination of K^+ by N3' in the morpholino ring of mA30 forces it into metastable configurations (Fig. Ion-Backbone-justfig), as can also be seen on the scatterplots, around the equator (Figure ResultFINAL-SIMRNAMNA C). As this obvious abnormality is attributable to an odd occurrence of the simulation and not the FF itself, data of mA30 and m31 are excluded in Table HelicalParams, Backbone. This behaviour was not observed in the preliminary RNA::MNA simulations (Figure XNAtrajectories). Except for the χ dihedral angle with a deviation of 2.39-2.93°, no significant differences in average backbone angles in the RNA backbone could be observed in the RNA::MNA duplex versus the model obtained for dsRNA (Table Backbone). However, helical parameters significantly deviate from the A-type duplex (Table HelicalParams) with a deeper major groove, and conversely more narrow minor groove, an increase in basepairs per turn and therefore helical size, along with a larger diameter. To investigate a possible correlation of the altered χ angle in the RNA strand to the overall shape for nucleic acid duplex, we performed two supplementary RNA::RNA double-DDD simulations (Figure ChiangleTrajPucker, ChiangleCurves). While in the original simulations the χ torsion angle was unrestrained, in the additional simulations restraints are applied on χ torsion angles at 194° and 189° respectively to mimic the impact on the overall helix structure by increasing deviation from the 199° χ -angle in standard A-type duplexes, at 5° increments. The helical parameters of the RNA::RNA duplex with χ torsion angles restraints at 194° are in line to those of the unrestrained RNA::MNA duplex (Table HelicalParams, ChiBackbone). The observed deviations from the A-type duplex are more pronounced in RNA::RNA duplex with χ torsion angles restraints at 189°.

To compare the orientation of P atoms in the linker in complementary strands, the dot product of the cross product of the OP1-P-OP2 plane of the leading RNA nucleotide and complementary MNA

nucleotide was calculated and compared to values obtained for corresponding base pairs in 3 simulations performed on dsRNA (Figure LinkerOrientation). Normalised distribution per duplex type shows a clear peak for the RNA::RNA duplexes around 137°–139° while for RNA::MNA it peaks at 120°–122°. An overlay of a set of the unrestrained RNA::RNA and RNA::MNA C-G basepairs, shows that not only the orientation of P in the MNA strand is changed but also its position relative to the corresponding P in dsRNA.

3.4. Discussion

Xeno-nucleic acids (XNAs) are a class of synthetic molecules that differ from naturally occurring nucleic acids in the use of altered backbone that can impact the three dimensional structure, stability and processing by naturally occurring biological processes. Results on numerous XNAs have been synthesized and characterized in the last decades revealed that ability to form stable homoduplexes and heteroduplexes with natural nucleic acids determines potential applications [24]. Molecular modeling simulations can provide understanding and predict how changes in the sugar-phosphate backbone impact on the complementation properties of the nucleic acids. However, the missing link in this field is a software tool that builds initial XNA models beyond ribose-based nucleic acids and robust force field parametrization to perform molecular mechanics simulations within standard packages such as AMBER. To fill this gap, we introduced a streamlined molecular modeling approach starting from quantum mechanics on XNA fragments. The versatile nucleic acid builder (Ducque) presented here builds XNA structures with unlimited backbone chemistry. It relies on a library of nucleosides and linkers that is user-extendable. We demonstrated the tool can build initial models starting from preferably low energy conformations of nucleosides and linkers, extracted from conformational sampling experiments. The latter also delivered data for force field parametrisation required to perform molecular mechanics simulations on the initial model. Ducque produced models of RNA and DNA duplexes as fast as the NAB tool. The latter is part of the AMBER toolkit and requires other AMBER programs (like tLEaP) to run, whereas Ducque functions as a standalone software. Syntactically, Ducque reads in simple text files to directly generate a model, as opposed to NAB's DSL which writes like C code and requires compilation of and executing the produced binary. Produced initial models are highly similar for both model builders and provided the same results in a subsequent MD simulation in standard force fields of AMBER. In contrast to NAB, Ducque is not limited to ribose based nucleic acids. Low energy conformations for non-ribose based nucleosides with hexitol and xylose sugars were included in the Ducque library, among others, and served to build nucleic acid duplexes for which MD simulations could be performed using a parametrised force fields within AMBER [49, 8]. As sugar puckering is crucial for the backbone geometry and flexibility, we used Paramfit to optimise dihedral force field parameters on non-ribose nucleosides by fitting QM to MM energies of selected conformations of nucleosides after RESP charges had been calculated through the ORCA implementation. Linker parameters from GAFF2 were validated through QM approaches.

The initial RNA::HNA duplex was built and remained stable during the MD round. Ducque's model of dXyNA homoduplex with backbone dihedrals for a ladder-like structure converged towards the left-handed helical structure in the MD simulation, confirming what was predicted the dXyNA duplex (62, 63, 64) (Figure S13). This result, and that of the DNA::RNA heteroduplex, ensured that dihedral angles used for model building in Ducque did not determine the outcome of the MD simulation.

The methodology described to derive charges for naturally occurring RNA modifications [41] starts off differently than the one described in this work, but has the same goal. At last, experiments were performed on an RNA::MNA duplex. The charges for the MNA chemistry were derived using the new ORCA implementation, that was evaluated on HNA. According to the proposed workflow, low energy conformations of nucleosides and linker are described by QM calculations and ported to the Ducque library. In line with the available crystal structure [69], the 4C_1 chair having the nucleobase in an equatorial position turned out to be the lowest energy conformation of morpholino nucleosides according to QM. An initial model was produced by ducque and subjected to an mm simulation using a force field that was parametrised as described in the methods section. the rna::mna heteroduplex simulations yielded a right-handed helical structure that does not belong to the A- or B-type family [70]. The morpholino rings remain predominantly in their chair conformation with an equatorial orientation of the nucleobase. Also in homoduplexes of $(4' \rightarrow 6')-(\beta\text{-D-glucopyranosyl})$ oligonucleotides (or β -homo-DNA) [3], $(4' \rightarrow 6')$ and $(3' \rightarrow 6')$ pentopyranose systems [71] this conformation is adopted by the six-membered pyranoses in the backbone but none of these XNA chemistries is able to communicate by base-pairing with either RNA or DNA. Changing the C1' configuraton (e.g. α -homo-DNA) or having the nucleobase on a C2' axial position instead of the regular C1' anomeric (equatorial) site of the six membered ring in the 'glycon' moiety (HNA), generated XNA systems that cross-pair with natural nucleic acids [6].

Except for a Watson-Crick-Franklin (WCF) to Hoogsteen (HG) transition in a mA:U basepair (Fig. ref{fig:ionbackbone-justfig"}, biblio)), that was facilitated by K⁺ binding and fraying at the helix end, stable WCF base pairing was observed during the MM simulation. Transitions between WCF and HG could be important in DNA recognition and replication, but are difficult to investigate experimentally because they are rare and short-lived. A transition from WCF to HG basepairing had been observed before for DNA in relaxation dispersion NMR experiments [72]. The same outside-route transition from WCF to HG base pairing was observed in DNA simulations, as the nucleotide temporarily unpairs and interacts with the solute, before repairing [73]. The occurrence of a WCF to HG transition in the final MNA::RNA simulation demonstrates that our optimised dihedral force field parameters allow puckering in the morpholino ring that is also expected in real life.

Only subtle changes in dihedral angles of the RNA strand, compared to values measured in dsRNA A-type helices, are imposed by the complementary MNA. To complement MNA, RNA adapts to the conformational constraints of the rigid binding partner with a more *anti*-characterised χ torsion angles while other dihedral angles remain closely to what is observed in dsRNA (Table Backbone, Fig. ChiangleCurves). This minor difference significantly changes helical parameters for helicity, groove widths, roll and twist of the duplex, as was demonstrated by restrained MD on dsRNA resulted in the same effect that is more pronounced if restraints on χ torsion angles deviate further from the 199° observed in A-type duplexes (Table HelicalParams). The increased diameter can be attributed to the six-membered morpholino ring in the backbone, in line to other XNA with six-membered sugar moieties [65, 74]. The unique chemistry of the MNA backbone changed orientation of the phosphoramidate in MNA and displaces the P atom in a position where it does not align with phosphates in the backbone of standard RNA (Fig. LinkerOrientation). Combined with the significantly different helical parameters, like increased radius of the double helix, altered positioning of the phosphoramidate due to the morpholino ring, can all contribute to the lack of RNA degradation when bound to complementary MNA sequences [15].

3.5. Conclusion

To conclude, the current work describes Ducque, a free and open-source program to construct initial models for (synthetic) nucleic acid duplexes of virtually any XNA chemistry. The manuscript proves that Ducque easily builds any type of structure for a given chemistry and that, with the correct force field, an accurate prediction can be modelled. We demonstrated one can extend the nucleic acid library with new sugars and nucleoside linkers. Ducque requires only built-in and NumPy modules, and Tkinter for the GUI. The user can supply Ducque with chemistries and their respective conformations, which can be confidently generated for with the conformational sampling scheme when no experimental structure is available. The current work describes Ducque, a free and open-source program to construct initials models for (synthetic) nucleic acid duplexes of virtually any XNA chemistry. The DNA and RNA chemistries were already well documented in literature with both crystallographic and NMR structures. This allowed parameters to be fitted on nucleosides whose behaviour was readily described in detail. For a given XNA that has not benefitted yet from years of characterisation, it is shown that the computed PESs are exceptionally accurate in predicting its behaviour.

The manuscript proves that not only can Ducque easily build any type of structure for a given chemistry, but that with the correct force field an accurate prediction can be modelled. We demonstrated one can extend the nucleic acid library with new sugars and nucleoside linkers. Ducque requires only built-in and NumPy modules, and Tkinter for the GUI. The user can supply Ducque with chemistries and their respective conformations, which can be confidently generated for with the conformational sampling scheme. On top of that, a suitable force field can be derived.

The RNA::MNA heteroduplex simulations yielded a right-handed helical structure that does not belong to the A- or B-type family. Only subtle changes in dihedral angles of the RNA strand, compared to values measured in dsRNA A-type helices, are imposed by the complementary MNA. The significantly different helical parameters, like increased radius of the double helix and altered positioning of the phosphoramidate due to the morpholino ring, can all contribute to the lack of RNA degradation when bound to complementary MNA sequences. Except for a Watson-Crick-Franklin to Hoogsteen transition in a mA:U basepair, that was facilitated by K⁺ binding and fraying at the helix end, stable WCF base pairing was observed during the MM simulation. This corroborates well with the available data that the heteroduplex differs from either A- and B-type duplexes and it is strongly suspected it acts as a steric block, in the ASO pathways, as it has also been demonstrated the heteroduplex is not fit for degradation by the RNase H enzyme [15, 70].

3.5.1. Supplementary information

Supplementary Data is available at NAR Online. The Ducque GitHub repository contains an extensive and comprehensible manual on how to start with Ducque and broaden the library of chemistries (Ducque/docs/). It also contains a separate guide on how to build a force field (Ducque/ff/). The latter directory contains the Morpholino FF used in this project.

Ducque is available on GitHub github.com/jrihon/Ducque and works on most Linux operating system (tested on Ubuntu 20 and 22 LTS), it runs on MacOS and *Windows Subsystem for Linux* (WSL).

CeNA, β -homo DNA, (d)XyNA, 2'-O-Me RNA, 2'-Fluoro RNA, HNA and MNA are included. It produces full homoduplexes, heteroduplexes and allows mixed chemistries in the complementary strand. The

leading sequence is also allowed to vary in chemistry, though we advise to keep the leading strand within the same XNA type.

A separate guide is provided on how to build a force field, how AMBER interprets its atoms, terminology, requirements for the creation of force fields, on how to perform parametrisation (resp, equivalence, restraints, paramfit, etc.) on a local machine. All figures are generated through python-scripted methods (Libraries : matplotlib, cartopy, NumPy, SciPy, mayavi) and/or concatenated/adapted in InkScape. All python libraries used are freely available through conda and/or pip (NumPy, SciPy, Tkinter, Cartopy, Mayavi).

3.6. References

- [1] Gait. M., Agrawal. S., "Introduction and History of the Chemistry of Nucleic Acids Therapeutics" *Antisense RNA Design, Delivery, and Analysis*, (NO VOLUME), pp. 3, 2022, doi: 10.1007/978-1-0716-2010-6_1
- [2] Chaput. J., Herdewijn. P., "What Is XNA?" *Angewandte Chemie International Edition*, vol. 58, pp. 11570, 2019-07, doi: 10.1002/anie.201905999
- [3] Egli. M., Manoharan. M., "Chemistry, structure and function of approved oligonucleotide therapeutics" *Nucleic Acids Research*, vol. 51, pp. 2529, 2023-03, doi: 10.1093/nar/gkad067
- [4] Pinheiro. V., Taylor. A., Cozens. C., Abramov. M., Renders. M., Zhang. S., Chaput. J., Wengel. J., Peak-Chew. S., McLaughlin. S., Herdewijn. P., Holliger. P., "Synthetic Genetic Polymers Capable of Heredity and Evolution" *Science*, vol. 336, pp. 341, 2012-04, doi: 10.1126/science.1217622
- [5] Groaz. E., Herdewijn. P., "Hexitol Nucleic Acid (HNA): From Chemical Design to Functional Genetic Polymer" *Handbook of Chemical Biology of Nucleic Acids*, (NO VOLUME), pp. 1, 2023, doi: 10.1007/978-981-16-1313-5_15-1
- [6] Lescrinier. E., Froeyen. M., Herdewijn. P., "Difference in conformational diversity between nucleic acids with a six-membered 'sugar' unit and natural 'furanose' nucleic acids" *Nucleic acids research*, vol. 31, pp. 2975–2989, 2003, NO DOI
- [7] Koshkin. A., Singh. S., Nielsen. P., Rajwanshi. V., Kumar. R., Meldgaard. M., Olsen. C., Wengel. J., "LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bi-cyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition" *Tetrahedron*, vol. 54, pp. 3607, 1998-04, doi: 10.1016/s0040-4020(98)00094-5
- [8] Mattelaer. C., Maiti. M., Smets. L., Maiti. M., Schepers. G., Mattelaer. H., Rosemeyer. H., Herdewijn. P., Lescrinier. E., "Stable Hairpin Structures Formed by Xylose-Based Nucleic Acids" *ChemBioChem*, vol. 22, pp. 1638, 2021-02, doi: 10.1002/cbic.202000803
- [9] Yang. H., Eremeeva. E., Abramov. M., Jacquemyn. M., Groaz. E., Daelemans. D., Herdewijn. P., "CRISPR-Cas9 recognition of enzymatically synthesized base-modified nucleic acids" *Nucleic Acids Research*, vol. 51, pp. 1501, 2023-01, doi: 10.1093/nar/gkac1147
- [10] Roshmi. R., Yokota. T., "Viltolarsen for the treatment of Duchenne muscular dystrophy" *Drugs of Today*, vol. 55, pp. 627, 2019, doi: 10.1358/dot.2019.55.10.3045038
- [11] Chen. S., Le. B., Rahimizadeh. K., Shaikh. K., Mohal. N., Veedu. R., "Synthesis of a Morpholino Nucleic Acid (MNA)-Uridine Phosphoramidite, and Exon Skipping Using MNA/2'-O-Methyl Mixmer Antisense Oligonucleotide" *Molecules*, vol. 21, pp. 1582, 2016-11, doi: 10.3390/molecules21111582
- [12] Moulton. J. "Using Morpholinos to Control Gene Expression" *Current Protocols in Nucleic Acid Chemistry*, vol. 68, NO PAGERANGE2017-03, doi: 10.1002/cpnc.21
- [13] Corey. D., Abrams. J., NO TITLE, *Genome Biology*, vol. 2, pp. reviews1015.1, 2001, doi: 10.1186/gb-2001-2-5-reviews1015
- [14] Roberts. T., Langer. R., Wood. M., "Advances in oligonucleotide drug delivery" *Nature Reviews Drug Discovery*, vol. 19, pp. 673, 2020-08, doi: 10.1038/s41573-020-0075-7
- [15] Nan. Y., Zhang. Y., "Antisense Phosphorodiamidate Morpholino Oligomers as Novel Antiviral Compounds" *Frontiers in Microbiology*, vol. 9, NO PAGERANGE2018-04, doi: 10.3389/fmicb.2018.00750
- [16] Lebleu. B., Moulton. H., Abes. R., Ivanova. G., Abes. S., Stein. D., Iversen. P., Arzumanov. A., Gait. M., "Cell penetrating peptide conjugates of steric block oligonucleotides" *Advanced Drug Delivery Reviews*, vol. 60, pp. 517, 2008-03, doi: 10.1016/j.addr.2007.09.002
- [17] Vanmeert. M., Razzokov. J., Mirza. M., Weeks. S., Schepers. G., Bogaerts. A., Rozenski. J., Froeyen. M., Herdewijn. P., Pinheiro. V., Lescrinier. E., "Rational design of an XNA ligase through docking of unbound nu-

- cleic acids to toroidal proteins" *Nucleic Acids Research*, vol. 47, pp. 7130, 2019-06, doi: 10.1093/nar/gkz551
- [18] Macke. T. "NAB: a language for molecular manipulation" NO PARENT FOUND, NO PARENT FOUND, NO PAGERANGE1996, NO DOI
- [19] Alenaizan. A., Barnett. J., Hud. N., Sherrill. C., Petrov. A., "The proto-Nucleic Acid Builder: a software tool for constructing nucleic acid analogs" *Nucleic Acids Research*, vol. 49, pp. 79, 2020-12, doi: 10.1093/nar/gkaa1159
- [20] Lu. X., Olson. W., "3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures" *Nature Protocols*, vol. 3, pp. 1213, 2008-07, doi: 10.1038/nprot.2008.104
- [21] Lavery. R., Moakher. M., Maddocks. J., Petkeviciute. D., Zakrzewska. K., "Conformational analysis of nucleic acids revisited: Curves+" *Nucleic Acids Research*, vol. 37, pp. 5917, 2009-07, doi: 10.1093/nar/gkp608
- [22] Duffy. K., Arangundy-Franklin. S., Holliger. P., "Modified nucleic acids: replication, evolution, and next-generation therapeutics" *BMC Biology*, vol. 18, NO PAGERANGE2020-09, doi: 10.1186/s12915-020-00803-6
- [23] Eremeeva. E., Herdewijn. P., "Non canonical genetic material" *Current Opinion in Biotechnology*, vol. 57, pp. 25, 2019-06, doi: 10.1016/j.copbio.2018.12.001
- [24] Anosova. I., Kowal. E., Dunn. M., Chaput. J., Van Horn. W., Egli. M., "The structural diversity of artificial genetic polymers" *Nucleic Acids Research*, vol. 44, pp. 1007, 2015-12, doi: 10.1093/nar/gkv1472
- [25] Mattelaer. C., Mattelaer. H., Rihon. J., Froeyen. M., Lescrinier. E., "Efficient and Accurate Potential Energy Surfaces of Puckering in Sugar-Modified Nucleosides" *Journal of Chemical Theory and Computation*, vol. 17, pp. 3814, 2021-05, doi: 10.1021/acs.jctc.1c00270
- [26] Deserno. M. "How to generate equidistributed points on the surface of a sphere" *If Polymerforshung* (Ed.), vol. 99, NO PAGERANGE2004, NO DOI
- [27] Cremer. D., Pople. J., "General definition of ring puckering coordinates" *Journal of the American Chemical Society*, vol. 97, pp. 1354, 1975-03, doi: 10.1021/ja00839a011
- [28] Strauss. H., Pickett. H., "Conformational structure, energy, and inversion rates of cyclohexane and some related oxanes" *Journal of the American Chemical Society*, vol. 92, pp. 7281, 1970-12, doi: 10.1021/ja00728a009
- [29] Neese. F., Wennmohs. F., Becker. U., Ripplinger. C., "The ORCA quantum chemistry program package" *The Journal of Chemical Physics*, vol. 152, NO PAGERANGE2020-06, doi: 10.1063/5.0004608
- [30] Harris. C., Millman. K., Gommers. R., Virtanen. P., Cournapeau. D., Wieser. E., Taylor. J., Berg. S., Smith. N., Kern. R., Picus. M., Hoyer. S., Brett. M., Haldane. A., Wiebe. M., Peterson. P., Gérard-Marchant. P., Shepard. K., Reddy. T., Weckesser. W., Abbasi. H., Gohlke. C., Oliphant. T., "Array programming with NumPy" *Nature*, vol. 585, pp. 357, 2020-09, doi: 10.1038/s41586-020-2649-2
- [31] Lavery. R., Moakher. M., Maddocks. J., Petkeviciute. D., Zakrzewska. K., "Conformational analysis of nucleic acids revisited: Curves+" *Nucleic Acids Research*, vol. 37, pp. 5917, 2009-07, doi: 10.1093/nar/gkp608
- [32] Case. D., Cheatham. T., Darden. T., Gohlke. H., Luo. R., Merz. K., Onufriev. A., Simmerling. C., Wang. B., Woods. R., "The Amber biomolecular simulation programs" *Journal of Computational Chemistry*, vol. 26, pp. 1668, 2005-09, doi: 10.1002/jcc.20290
- [33] Salomon-Ferrer. R., Case. D., Walker. R., "An overview of the Amber biomolecular simulation package" *WIREs Computational Molecular Science*, vol. 3, pp. 198, 2012-09, doi: 10.1002/wcms.1121
- [34] Bayly. C., Cieplak. P., Cornell. W., Kollman. P., "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model" *The Journal of Physical Chemistry*, vol. 97, pp. 10269, 1993-10, doi: 10.1021/j100142a004
- [35] Singh. U., Kollman. P., "An approach to computing electrostatic charges for molecules" *Journal of Com-*

- putational Chemistry*, vol. 5, pp. 129, 1984-04, doi: 10.1002/jcc.540050204
- [36] Besler. B., Merz. K., Kollman. P., "Atomic charges derived from semiempirical methods" *Journal of Computational Chemistry*, vol. 11, pp. 431, 1990-05, doi: 10.1002/jcc.540110404
- [37] Dupradeau. F., Pigache. A., Zaffran. T., Savineau. C., Lelong. R., Grivel. N., Lelong. D., Rosanski. W., Cieplak. P., "The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building" *Physical Chemistry Chemical Physics*, vol. 12, pp. 7821, 2010, doi: 10.1039/c0cp00111b
- [38] Cieplak. P., Cornell. W., Bayly. C., Kollman. P., "Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins" *Journal of Computational Chemistry*, vol. 16, pp. 1357, 1995-11, doi: 10.1002/jcc.540161106
- [39] Sanner. M., Olson. A., Spehner. J., "Reduced surface: An efficient way to compute molecular surfaces" *Biopolymers*, vol. 38, pp. 305, 1996-03, doi: 10.1002/(sici)1097-0282(199603)38:3<305::aid-bip4>3.0.co;2-y
- [40] Connolly. M. "Analytical molecular surface calculation" *Journal of Applied Crystallography*, vol. 16, pp. 548, 1983-10, doi: 10.1107/s0021889883010985
- [41] Aduri. R., Psciuk. B., Saro. P., Taniga. H., Schlegel. H., SantaLucia. J., "AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA" *Journal of Chemical Theory and Computation*, vol. 3, pp. 1464, 2007-06, doi: 10.1021/ct600329w
- [42] Chirlian. L., Francl. M., "Atomic charges derived from electrostatic potentials: A detailed study" *Journal of Computational Chemistry*, vol. 8, pp. 894, 1987-09, doi: 10.1002/jcc.540080616
- [43] Barca. G., Bertoni. C., Carrington. L., Datta. D., De Silva. N., Deustua. J., Fedorov. D., Gour. J., Gunina. A., Guidez. E., Harville. T., Irle. S., Ivanic. J., Kowalski. K., Leang. S., Li. H., Li. W., Lutz. J., Magoulas. I., Mato. J., Mironov. V., Nakata. H., Pham. B., Piecuch. P., Poole. D., Pruitt. S., Rendell. A., Roskop. L., Ruedenberg. K., Sat-sasathuchana. T., Schmidt. M., Shen. J., Slipchenko. L., Sosonkina. M., Sundriyal. V., Tiwari. A., Galvez Vallejo. J., Westheimer. B., Włoch. M., Xu. P., Zahariev. F., Gordon. M., "Recent developments in the general atomic and molecular electronic structure system" *The Journal of Chemical Physics*, vol. 152, NO PAGERANGE2020-04, doi: 10.1063/5.0005188
- [44] Betz. R., Walker. R., "Paramfit: Automated optimization of force field parameters for molecular dynamics simulations" *Journal of Computational Chemistry*, vol. 36, pp. 79, 2014-11, doi: 10.1002/jcc.23775
- [45] Weiner. S., Kollman. P., Case. D., Singh. U., Ghio. C., Alagona. G., Profeta. S., Weiner. P., "A new force field for molecular mechanical simulation of nucleic acids and proteins" *Journal of the American Chemical Society*, vol. 106, pp. 765, 1984-02, doi: 10.1021/ja00315a051
- [46] Darden. T., York. D., Pedersen. L., "Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems" *The Journal of Chemical Physics*, vol. 98, pp. 10089, 1993-06, doi: 10.1063/1.464397
- [47] Zgarbová. M., Šponer. J., Otyepka. M., Cheatham. T., Galindo-Murillo. R., Jurečka. P., "Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA" *Journal of Chemical Theory and Computation*, vol. 11, pp. 5723, 2015-11, doi: 10.1021/acs.jctc.5b00716
- [48] Zgarbová. M., Otyepka. M., Šponer. J., Mládek. A., Banáš. P., Cheatham. T., Jurečka. P., "Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles" *Journal of Chemical Theory and Computation*, vol. 7, pp. 2886, 2011-08, doi: 10.1021/ct200162x
- [49] Schofield. P., Taylor. A., Rihon. J., Peña Martinez. C., Zinn. S., Mattelaer. C., Jackson. J., Dhaliwal. G., Schepers. G., Herdewijn. P., Lescrinier. E., Christ. D., Holliger. P., "Characterization of an HNA aptamer suggests a non-canonical G-quadruplex motif" *Nucleic Acids Research*, vol. 51, pp. 7736, 2023-07, doi: 10.1093/nar/gkad592

- [50] Kräutler. V., Hünenberger. P., "A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations" *Journal of Computational Chemistry*, vol. 22, pp. 501, 2001, doi: 10.1002/1096-987x(20010415)22:5<501::aid-jcc1021>3.0.co;2-v
- [51] Davidchack. R., Handel. R., Tretyakov. M., "Langevin thermostat for rigid body dynamics" *The Journal of Chemical Physics*, vol. 130, NO PAGERANGE2009-06, doi: 10.1063/1.3149788
- [52] Berendsen. H., Postma. J., DiNola. A., Haak. J., "Molecular dynamics with coupling to an external bath" *The Journal of Chemical Physics*, vol. 81, pp. 3684, 1984-10, doi: 10.1063/1.448118
- [53] Roe. D., Cheatham. T., "PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data" *Journal of Chemical Theory and Computation*, vol. 9, pp. 3084, 2013-06, doi: 10.1021/ct400341p
- [54] Salomon-Ferrer. R., Götz. A., Poole. D., Le Grand. S., Walker. R., "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald" *Journal of Chemical Theory and Computation*, vol. 9, pp. 3878, 2013-08, doi: 10.1021/ct400314y
- [55] Le Grand. S., Götz. A., Walker. R., "SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations" *Computer Physics Communications*, vol. 184, pp. 374, 2013-02, doi: 10.1016/j.cpc.2012.09.022
- [56] Humphrey. W., Dalke. A., Schulten. K., "VMD: Visual molecular dynamics" *Journal of Molecular Graphics*, vol. 14, pp. 33, 1996-02, doi: 10.1016/0263-7855(96)00018-5
- [57] Pettersen. E., Goddard. T., Huang. C., Couch. G., Greenblatt. D., Meng. E., Ferrin. T., "UCSF Chimera—A visualization system for exploratory research and analysis" *Journal of Computational Chemistry*, vol. 25, pp. 1605, 2004-07, doi: 10.1002/jcc.20084
- [58] Vanquelef. E., Simon. S., Marquant. G., Garcia. E., Klimerak. G., Delepine. J., Cieplak. P., Dupradeau. F., "R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments" *Nucleic Acids Research*, vol. 39, pp. W511–W517, 2011-05, doi: 10.1093/nar/gkr288
- [59] Sychrovsky. V., Foldynova-Trantirkova. S., Spackova. N., Robeyns. K., Van Meervelt. L., Blankenfeldt. W., Vokacova. Z., Sponer. J., Trantirek. L., "Revisiting the planarity of nucleic acid bases: Pyramidilization at glycosidic nitrogen in purine bases is modulated by orientation of glycosidic torsion" *Nucleic Acids Research*, vol. 37, pp. 7321, 2009-09, doi: 10.1093/nar/gkp783
- [60] Zierkiewicz. W., Komorowski. L., Michalska. D., Cerny. J., Hobza. P., "The Amino Group in Adenine: MP2 and CCSD(T) Complete Basis Set Limit Calculations of the Planarization Barrier and DFT/B3LYP Study of the Anharmonic Frequencies of Adenine" *The Journal of Physical Chemistry B*, vol. 112, pp. 16734, 2008-12, doi: 10.1021/jp8058118
- [61] Davis. R., Shaban. N., Perrino. F., Hollis. T., "Crystal structure of RNA-DNA duplex provides insight into conformational changes induced by RNase H binding" *Cell Cycle*, vol. 14, pp. 668, 2015-02, doi: 10.4161/15384101.2014.994996
- [62] Maiti. M., Siegmund. V., Abramov. M., Lescrinier. E., Rosemeyer. H., Froeyen. M., Ramaswamy. A., Ceulemans. A., Marx. A., Herdewijn. P., "Solution Structure and Conformational Dynamics of Deoxyxylonucleic Acids (dXNA): An Orthogonal Nucleic Acid Candidate" *Chemistry – A European Journal*, vol. 18, pp. 869, 2011-12, doi: 10.1002/chem.201102509
- [63] Ramaswamy. A., Froeyen. M., Herdewijn. P., Ceulemans. A., "Helical Structure of Xylose-DNA" *Journal of the American Chemical Society*, vol. 132, pp. 587, 2009-12, doi: 10.1021/ja9065877
- [64] Ramaswamy. A., Smyrnova. D., Froeyen. M., Maiti. M., Herdewijn. P., Ceulemans. A., "Molecular Dynamics of Double Stranded Xylo-Nucleic Acid" *Journal of Chemical Theory and Computation*, vol. 13, pp.

5028, 2017-09, doi: 10.1021/acs.jctc.7b00309

- [65] Lescrinier. E., Esnouf. R., Schraml. J., Busson. R., Heus. H., Hilbers. C., Herdewijn. P., "Solution structure of a HNA–RNA hybrid" *Chemistry & Biology*, vol. 7, pp. 719, 2000-09, doi: 10.1016/s1074-5521(00)00017-x
- [66] Wang. J., Wolf. R., Caldwell. J., Kollman. P., Case. D., "Development and testing of a general amber force field" *Journal of Computational Chemistry*, vol. 25, pp. 1157, 2004-04, doi: 10.1002/jcc.20035
- [67] Zhang. C., Lu. C., Wang. Q., Ponder. J., Ren. P., "Polarizable Multipole-Based Force Field for Dimethyl and Trimethyl Phosphate" *Journal of Chemical Theory and Computation*, vol. 11, pp. 5326, 2015-10, doi: 10.1021/acs.jctc.5b00562
- [68] IUPAC-IUB. J. "Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains" *European Journal of Biochemistry*, vol. 131, pp. 9–15, 1983, NO DOI
- [69] Zhang. W., Pal. A., Ricardo. A., Szostak. J., "Template-Directed Nonenzymatic Primer Extension Using 2-Methylimidazole-Activated Morpholino Derivatives of Guanosine and Cytidine" *Journal of the American Chemical Society*, vol. 141, pp. 12159, 2019-07, doi: 10.1021/jacs.9b06453
- [70] Langner. H., Jastrzebska. K., Caruthers. M., "Synthesis and Characterization of Thiophosphoramidate Morpholino Oligonucleotides and Chimeras" *Journal of the American Chemical Society*, vol. 142, pp. 16240, 2020-08, doi: 10.1021/jacs.0c04335
- [71] Eschenmoser. A. "Searching for Nucleic Acid Alternatives" NO PARENT FOUND, NO PARENT FOUND, pp. 5, 2011-02, doi: 10.1002/9780470977873.ch1
- [72] Nikolova. E., Kim. E., Wise. A., O'Brien. P., Andrecioaei. I., Al-Hashimi. H., "Transient Hoogsteen base pairs in canonical duplex DNA" *Nature*, vol. 470, pp. 498, 2011-01, doi: 10.1038/nature09775
- [73] Vreede. J., Bolhuis. P., Swenson. D., "Atomistic insight into the kinetic pathways for Watson–Crick to Hoogsteen transitions in DNA" *Nucleic Acids Research*, vol. 47, pp. 11069, 2019-10, doi: 10.1093/nar/gkz837
- [74] Egli. M., Pallan. P., Pattanayek. R., Wilds. C., Lubini. P., Minasov. G., Dobler. M., Leumann. C., Eschenmoser. A., "Crystal Structure of Homo-DNA and Nature's Choice of Pentose over Hexose in the Genetic System" *Journal of the American Chemical Society*, vol. 128, pp. 10847, 2006-08, doi: 10.1021/ja062548x

CURVATURE TORSION ON SYNTHETIC NUCLEIC ACIDS

4

4.1. Introduction

4.1.1. Welcome to the thesis

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

› An addition to the text

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

4.2. Methods

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

4.3. Result & Discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

4.4. Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

STRUCTURE PREDICTION OF AN HNA APTAMER WITH A G-QUADRUPLEX MOTIF

5

Adapted from the following manuscript 10.1093/nar/gkad592 :

Schofield P., Taylor A.I., Rihon J., Martinez C.D.P., Zinn S., Mattelaer C.-A., Jackson J., Dhaliwal G., Schepers G., Herdwijn P., Lescrinier E., Christ D., Holliger P. “Characterization of an HNA aptamer suggests a non-canonical G-quadruplex motif.” Nucleic Acids Research (NAR).

5.1. Introduction

5.1.1. Welcome to the thesis

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

› An addition to the text

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

5.2. Methods

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

5.3. Result & Discussion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

5.4. Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

CONCLUDING REMARKS AND FUTURE PERSPECTIVE

6

6.1. Concluding remarks

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

› An addition to the text

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

6.2. Future perspectives

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

› An addition to the text

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut.

Published works

List of Publications

- [1] Lorem ipsum publications Lorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum
- [2] Lorem ipsum publications Lorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum
- [3] Lorem ipsum publications Lorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum
- [4] Lorem ipsum publications Lorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum
- [5] Lorem ipsum publications Lorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum publicationsLorem ipsum

Software releases

- [1] Ducque
- [2] pucke.rs
- [3] pucke.py

Scientific outreach

EMBO stuff

Declarations

Scientific acknowledgements

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distingue possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

Personal contributions

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliquam quaerat voluptatem. Ut enim aequo doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distingue possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et.

Conflict of Interest

The author declares no conflicts of interest or any financial interests.