Rodrigo Vargas Honorato

Development and implementation and of a coarse-grained Protein-DNA/RNA docking model

Brazil

2017

Rodrigo Vargas Honorato

Development and implementation and of a coarse-grained Protein-DNA/RNA docking model

Research project submission as requirement of FAPESP Research Fellowships Abroad (BPE) for scientists who have a permanent position in a research institution in the State of São Paulo and obtain a leave of absence to develop a research project in a foreign institution.

Centro Nacional de Pesquisa em Energia e Materiais - CNPEM

Laboratório Nacional de Biociências - LNBio

Laboratório de Biologia Computacional - LBC

Brazil

2017

Abstract

The computational modeling of protein structures, interactions and dynamics has been advancing steadily in the past decades. Nonetheless researchers still face challenges when modeling biomolecular systems with large conformational spaces. Lowering the level of representation from all-atom to coarse-grained is a way to bypass limitations such as algorithmic efficiency and available computing power. The objective of a coarse-grained representation is to reduce the degrees of freedom, replacing side chains with pseudo-atoms. In order to develop coarse-grained force fields, efforts have been applied in two fronts; physics-based, following the same philosophy of it full-atom counterpart, basing it on molecular physics and knowledge-based that takes advantage the growing databases via statistical analysis. Here we will expound on the knowledge-based approach, the MARTINI force field. Four non-hydrogen atoms are mapped to one CG bead which describes one or more chemical building blocks along with its properties. Nucleotides are mapped to six or seven CG beads. The phosphate accounts for one bead and sugar for two, pyrimidines are represented as three-bead rings and purines as four-bead rings. Computational docking is a powerful tool to model three-dimensional structures of macromolecular interactions. By sampling a large number of possible conformations and selecting those with low interaction energies it is possible to obtain a native-like conformational of a macromolecular complex. The high ambiguity driven docking approach software HADDOCK developed in Alexandre Bonvin Lab uses biochemical and/or biophysical interactions to predict such interactions. The proposed research project aims to implement and benchmark a coarse-grained DNA/RNA model into HADDOCK. The use of coarse-grained models has enabled researchers to simulate large-scale biomolecular processes on time scales that were previously inaccessible to full-atom models. For such the following will be carried out; conversion of structures to CG models will be carried out by a conversion script adapted from the one provided by the research group responsible for MARTINI, adaptation of the MARTINI DNA/RNA topology and parameters to the CNS format, integration of the MARTINI CG DNA/RNA force field into HADDOCK, performance evaluation using a benchmark of Protein-DNA complexes and final assessment according to CAPRI criteria. Optimization of the scoring functions will consist of an evaluation of different ranges of weights in each of interaction energies and comparison with the CAPRI evaluation in order to obtain optimal weights to correctly identify the best generated complexes.

Keywords: Protein docking, macromolecular interaction, coarse-grained, MARTINI, HADDOCK

List of Figures

List of Tables

Contents

1 Introduction

1.1 Coarse grained models

The computational modeling of protein structures, interactions and dynamics has been advancing steadily in the past decades. Nonetheless researchers still face challenges when modeling biomolecular systems with large conformational spaces that require long simulation timescales. Lowering the level of representation from all-atom to coarse-grained is a way to bypass limitations such as algorithmic efficiency and available computing power (VENDRUSCOLO; DOBSON, 2011).

There are different scales of molecular modeling applications that can be described as a function of system size versus time scale (Figure 1). Methods based on all-atom representations usually cover up to the nanosecond timescale and are useful to observe mainly local motions. In terms of system size, simulations using all-atom representations are computationally feasible for systems that have at most, micro-molecular lengths. In order to study protein folding, global motions, aggregation or general dynamics of large molecular systems (up to milli-molecular length) one must take advantage of the more computationally effective coarse-grained representation that enables simulations at much longer time-scales for larger system sizes (KMIECIK et al., 2016).

The main objective of a coarse-grained representation is to reduce the number of degrees of freedom, replacing amino acids (either fragments or entire side chains) with

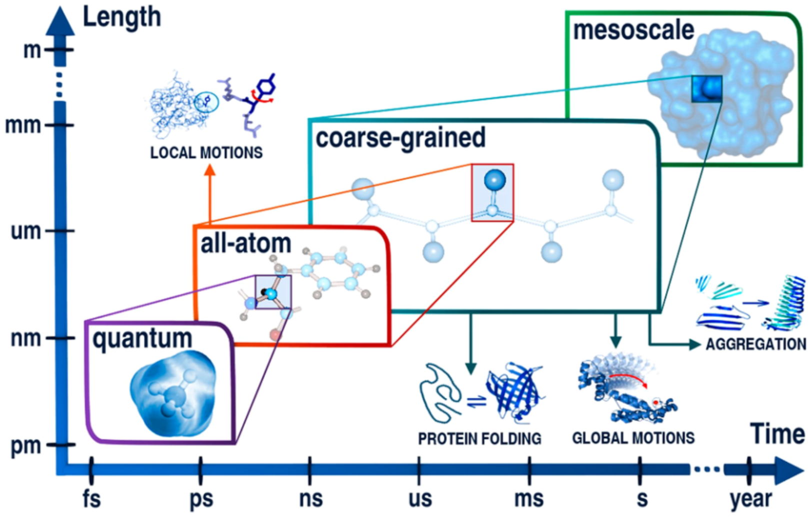


Figure 1 – Different scales of common computational biology experiments (KMIECIK et al., 2016).

pseudo-atoms. The first coarse-grained protein models were built in the 70's and were successfully able to simulate an entire folding process (LEVITT; WARSHEL, 1975). This early model consisted of pseudo atoms placed at $C\alpha$ positions and at the geometric center of protein side chains (except for glycine) and the only degree of freedom considered was the rotation angle along the central pseudo bond for three consecutive $C\alpha$ pseudo atoms. Interaction between united atoms was described as a simple Lennard-Jones potential.

Further development of the model accounted for the variable orientation of the pseudo sidechains and torsional potential for the main chain degrees of freedom were based on the statistical analysis of conformational properties of dipeptides (LEVITT, 1976). Similar models aimed to represent residue-residue interactions using statistical potentials and Monte Carlo to simulate the folding process (WILSON; DONIACH, 1989).

Most proteins adopt a well-defined three-dimensional arrangement defined by both the conformational restraints of the main-chain as well as interactions with the environment by the side chains, characterizing a convoluted interplay of intra and intermolecular bonds that defines its structure. Henceforth, it is expected that a coarse-grained representation of a protein contains all its crucial elements. The main objective of this representation is to reduce the number of degrees of freedom inherent to a protein structure.

The SICHO (side chain only) model considers only a pseudoatom placed in the center of each side chain, discarding backbone information; connectivity between pseudo atoms are given by statistically derived parameters observed in known structures (KOLINSKI et al., 1998). Novel methods such as The UNRES (united residue) (LIWO et al., 2014) and CABS (C-alpha, beta and side chain) (KURCINSKI et al., 2015) included information about protein's main chain as well as side chain. The Rosetta representation considers all backbone atoms and replaces the sidechain with a single pseudoatom (JOHNSON, 2004) (Figure 2).

The balance between enthalpy and entropy are affected by the reduction of the system's degrees of freedom. Since the impact in entropy is compensated by reducing the enthalpic terms, a coarse-grained model may accurately reproduce differences in free energy even though its enthalpy and entropy estimations might be inaccurate.

In order to develop coarse-grained force fields, efforts have been applied in two fronts; physics-based, following the same philosophy of it full-atom counterpart, basing it on molecular physics and knowledge-based that takes advantage the growing databases via statistical analysis. Here we will expound on the former, specifically on the MARTINI (MARRINK et al., 2007) force field.

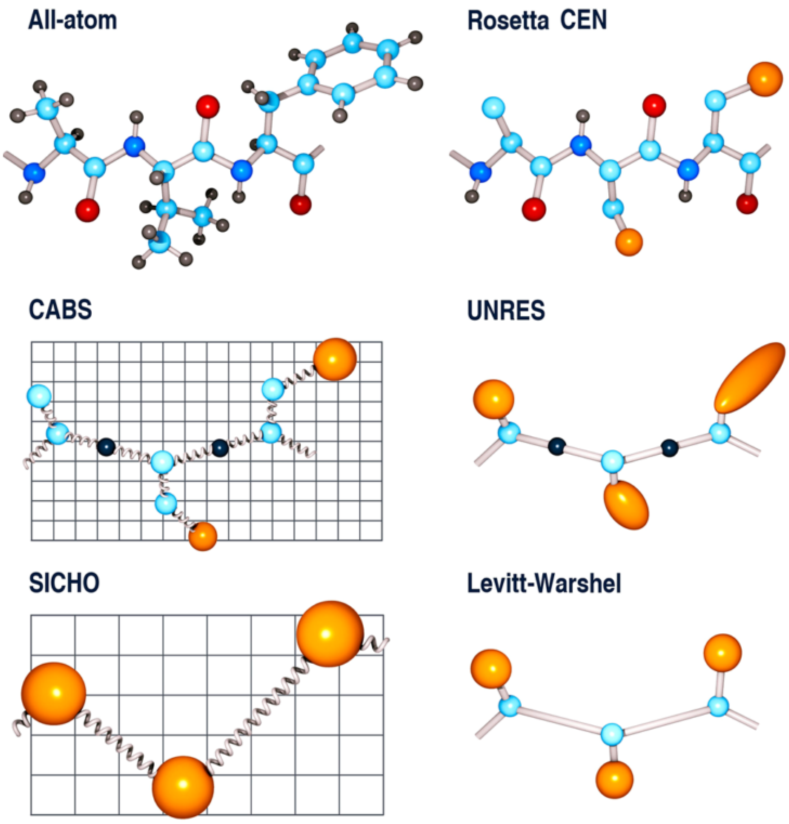


Figure 2 – Different coarse-grained representations of a protein; united side chain atoms are colored in orange. pseudobonds of fluctuating length are shown as springs and lattice models are shown on the underlying lattice slide (KMIECIK et al., 2016).

1.2 MARTINI Force Field

The MARTINI force field was developed taking into heavy consideration the atomistic models, especially for bonded interactions. It aims for a broader range of applications instead of focusing on accurate reproductions of structural details at a specific point in a given system. For that purpose, the non-bonded interactions of chemical building blocks are extensively calibrated against experimental data. The main focus of this coarse-graining approach is to offer a computationally optimized and user-friendly model that is applicable to a wide range of biologically relevant systems.

This model is based on a four-to-one mapping, four heavy atoms (and associated hydrogens) are represented by a single atom (CG bead). Four real water molecules are represented as a single CG bead whilst ion CG bead representation takes into account both the ion and its hydration shell (MARRINK; TIELEMAN, 2013) (Figure 3). The MARTINI model offers the following types of particles: charged (Q), apolar (C), non-polar (N) and polar (P); this is further expanded according to its hydrogen-bonding capabilities (donor, acceptor, both or none) or its degree of polarity $(1 - 5)$ giving a total of 18 types

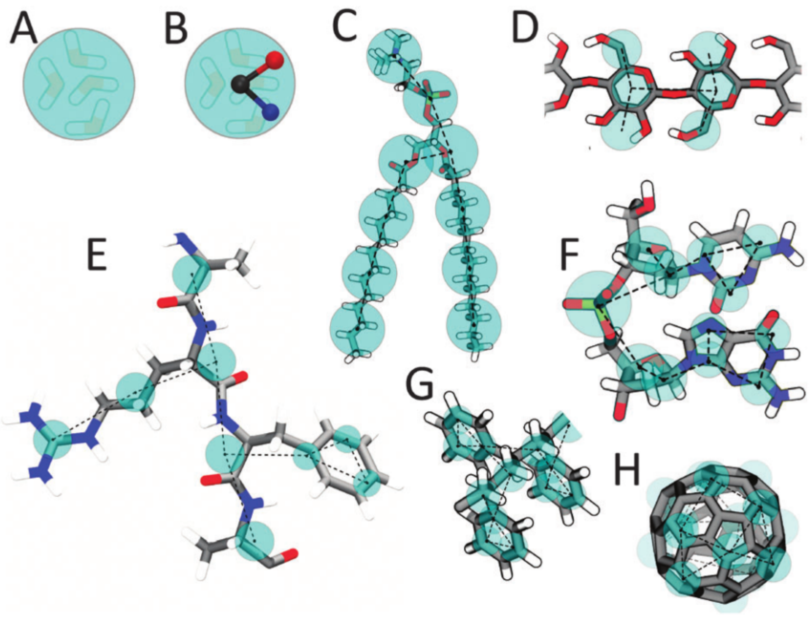


Figure 3 – Examples of Martini mapping selected molecules. (A) Standard water particle. (B) Polarizable water molecule with embedded charges. (C) DMPC lipid. (D) Polysaccharide fragment. (E) Peptide. (F) DNA fragment. (G) Polystyrene fragment. (H) Fullerene molecule. In all cases Martini CG beads are shown as cyan transparent beads overlaying the atomistic structure (MARRINK; TIELEMAN, 2013).

of particles.

Non-bonded interactions are described by a Lennard-Jones (LJ) 12-6 potential (equation 1.1), in which σ_{ij} represents the closest distance of approach between particles i, j and ϵ_{ij}, the strength of the interaction. Effective particle size is defined as $\sigma = 0.47nm$, if the given interaction is between charged (Q-type) and apolar (C-type) then $\sigma = 0.62nm$. This change in σ favours Q-type particles to keep their hydration shell when placed into an apolar medium (MARRINK et al., 2007).

$$E_{i,j}(r) = 4\epsilon \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^{6} \right] \tag{1.1}$$

Charged groups have a full charge q_{ij} that interacts via a Coulombic potential energy function (equation 1.2), with dielectric constant for explicit screening.

$$U_{el}(r) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1 r} \tag{1.2}$$

Bonded interactions are described by a weak harmonic potential $V_{bond}(R)$

$$V_{bond}(R) = \frac{1}{2} K_{bond} \left(R - R_{bond} \right)^2 \tag{1.3}$$

with equilibrium distance $R_{bond} = 0.47nm$ and a force constant of $K_{bond} = 1250kJmol^{-1}nm^{-2}$. To better represent the chemical structure of bonded interactions a weak harmonic potential $V_{angle}(\theta)$ is used

$$V_{angle}(\theta) = \frac{1}{2}K_{angle}\left\{cos(\theta) - cos(\theta_0)\right\}^2 \tag{1.4}$$

LJ interaction is not accounted for bonded particles but are calculated for second nearest neighbours (MARRINK et al., 2007).

1.3 MARTINI DNA Extension

The MARTINI CG DNA model was systematically parametrized according to the MARTINI philosophy, being backwards compatible with all other MARTINI model implementations. Experimental values such as liquid densities and partitioning free energies of small solutes between polar and nonpolar solvents are used to determine non-bonded interaction parameters(UUSITALO et al., 2015).

As previously described, four non-hydrogen atoms are mapped to one CG bead which describes one or more chemical building blocks along with its properties. The parametrization used by Uusitalo and collaborators combines top-down (experimental data) and bottom-up (atomistic simulations) methodologies to parametrize the CG DNA model. The CG bead types for each nucleobase were selected based on partition free energies from water to chloroform or hydrated octanol. Bonded interactions have been fitted to reproduce dihedral, angle and bond distributions from atomistic simulation of short single stranded DNAs (ssDNAs). For double stranded DNA (dsDNA) an elastic network was devised to maintain the double helical structure.

Here each nucleotide is mapped to six or seven CG beads. The phosphate accounts for one bead and sugar for two, pyrimidines are represented as three-bead rings and purines as four-bead rings (Figure 4). After changing the representation of dsDNA bases the resulting distance between them are 0.34 nm, which lead to issues in the CG model. To account for that, a smaller bead size was created for the CG DNA model with $\sigma = 0.32nm$ (68% smaller than the regular CG bead) and named T, for tiny.

Hydrogen bonding between bases is a crucial factor for the formation of dsDNA and since the standard MARTINI model does not describe the directional hydrogen bonds, the interaction between hydrogen bonding beads was tweaked. Interactions are defined in a pairwise fashion for each bead type, henceforth eight special beads were added for the purpose of replicating this crucial bond.

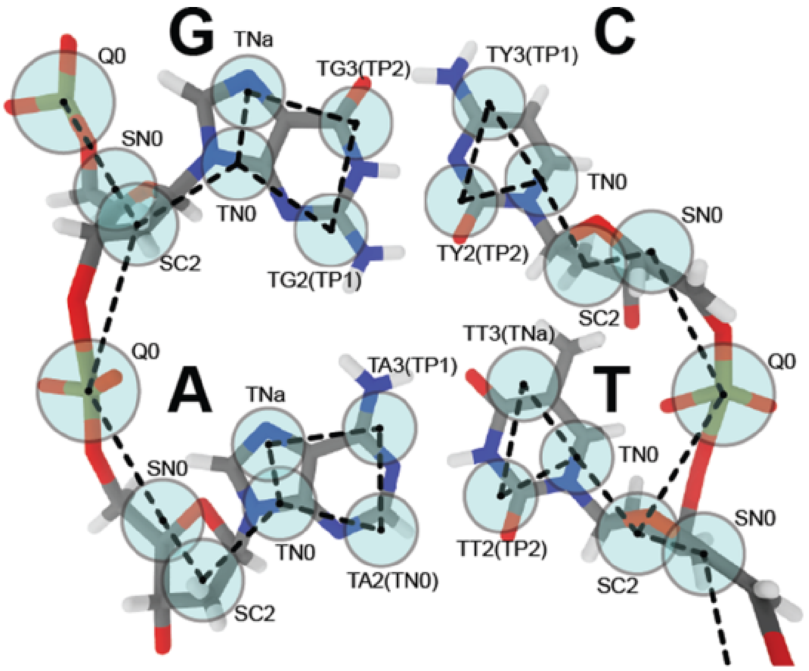


Figure 4 – DNA backbone is modeled with one bead describing the phosphate and two beads describing the sugar. The pyrimidines are modeled with three beads and the purines with four beads. T-prefix marks the beads that use the new tiny bead type. For hydrogen bonding beads, the new special bead types are shown together with the bead type describing their interactions with all beads except the special hydrogen bonding beads (in parentheses) (UUSITALO et al., 2015).

1.4 Integrative modelling of biomolecular complexes with HADDOCK

Knowledge of the atomistic, three-dimensional, interaction between macromolecules is fundamental for the understanding of biological functions of organisms, provided by its three-dimensional structure. These structures may be obtained using experimental methods such as X-Ray crystallography, nuclear magnetic resonance (NMR) or cryo electron microscopy (Cryo-EM). The RCSB Protein Data Bank (BERMAN, 2000) is the global repository for macromolecular structures with 125309 deposited structures (December/2016). Much can be studied and derived from experimentally solved complexes but the experimental obtention of three-dimensional complexes remain a timely, low throughput process.

Computational docking is a powerful tool to model three-dimensional structures of protein-protein, protein-DNA, protein-RNA and protein-small molecule interactions. By sampling a large number of possible conformations and selecting those with low interaction energies it is possible to obtain a native-like conformational of a macromolecular complex.

The high ambiguity driven docking approach software HADDOCK developed in Alexandre Bonvin Lab uses biochemical and/or biophysical interactions to predict macromolecular interactions and is renowned for its functionality and quality. Information on interacting residues was introduced as Ambiguous Interaction Restraints (AIRs) as means to drive the docking (DOMINGUEZ; BOELENS; BONVIN, 2003).

A worldwide experiment, the Critical Assessment of PRedicted Interactions (CAPRI) (JANIN, 2002) aims to evaluate the predictive power of docking software. HADDOCK was ranked in 2015 as the top software both in the Scorer and Server ranking categories (PERFORMANCE. . . , 2015).

1.4.1 Docking protocol

The docking approach is composed of python scripts derived from a software protocol called ARIA (Ambiguous Restraints for Iterative Assignment) (LINGE; ODONOGHUE; NILGES, 2001) used to integrate data from NRM into structure calculations and uses the highly flexible CNS (Crystallography & NMR System) (BRUNGER, 2013) software suit. Inter and intramolecular energies are accounted for using full electrostatic and van der Waals energy using the OPLS (JORGENSEN; TIRADO-RIVES, 1988) nonbonded parameters and further optimizations.

The docking protocol is executed in three steps: (1) rigid body energy minimizations after orientation randomization, *it0*; (2) semirigid simulated annealing in torsion angle space, *it1*; and (3) final refinement in Cartesian space with explicit solvent, *water*.

In the first step (*it0*), after two partner proteins are positioned at 150Å apart, each one is randomly rotated around its center of mass then submitted to a rigid body energy minimization. Firstly, four cycles of orientational optimization are executed so that each protein is allowed to rotate to minimize the intermolecular energy function, then macromolecules are allowed to both rotate and translate, thus performing the rigid body EM docking.

During the second stage (*it1*) there are three rounds of simulated annealing refinements; in the first one the two partners are considered rigid bodies and their orientation is optimized, in the second round the side chains at the interface are allowed to move and in the last both side chains and backbone (at the interface) are allowed to move in order to obtain the final conformational rearrangement.

On the last step of the docking protocol (*water*), the interacting system is placed in a 8Å shell of TIP3 water molecules (MARK; NILSSON, 2001), the system is heated to 300K with position restrains on all atoms except side chains. A short molecular dynamics is performed on the non-interface heavy atoms. On the final cooling stage the position restrains are limited to backbone atoms outside the interface. Final structures are then clustered using pairwise backbone interface root mean square deviation with

$(cutoff > 1.0\text{Å})$ and analyzed according to their average interaction energies E_{elec}, E_{vdw} and E_{acs} plus the average buried surface area.

2 Justification

Alexandre Bonvin's Lab is the Computational Structural Biology group at Utrecht University, located in the Netherlands. With high impact publications dating back to 1990, this research group is placed amongst the top ranking predictors and scorers on CAPRI. The use of experimental data to drive and/or validate its macromolecular predictions is one of the core methodologies used by this group and it yields exceptional results.

The proposed research project aims to implement and benchmark a coarse-grained DNA/RNA model into HADDOCK. The use of coarse-grained models has enabled researchers to simulate large-scale biomolecular processes on time scales that were previously inaccessible to full-atom models (INGÓLFSSON et al., 2013). Its usefulness has been proved on the fields of protein folding (LIWO et al., 1997), mechanosensitive channels (LOUHIVUORI et al., 2010), membrane protein self-assembly (PERIOLE et al., 2007) , DNA hybridization (SAMBRISKI; SCHWARTZ; PABLO, 2009) amongst many others.

This implementation will build upon a previous successful and thoroughly optimized implementation of the MARTINI Protein-Protein CG force field in HADDOCK (unpublished) and will facilitate computational biology studies with proteins and nucleic acids to accompany the ever growing scale and resolution of structural data.

As an example, the yeast RNA polymerase III elongation complex was determined at 3.9Å using Cryo-EM (HOFFMANN et al., 2015) consisting of 39284 atoms (without water), large systems such as this and the nucleossome with fibers varying from 10-nm to 30-nm (SILVA; OLIVEIRA; SANTOS, 2015), are the most adequate targets for CG models, granting deeper understanding of such relevant systems.

For such, a capable implementation done according to both the MARTINI philosophy and taking advantage HADDOCK's integrative approach to macromolecular interactions and its lab members competences, followed by a thorough benchmarking will lay the groundwork for profound studies of large scale macromolecular processes.

3 Objectives

3.1 General

Develop and implement a coarse-grained DNA/RNA model in HADDOCK and benchmark it for docking.

3.2 Specific

- Convert MARTINI coarse-grained DNA/RNA topology into CNS format (Crystallography & NMR System)
- Add and test forcefield in HADDOCK
- Convert Protein-DNA/RNA dataset do CG
- Run docking trials
- Evaluate results
- Optimize scoring functions
- Implement CG Protein-DNA/RNA docking

4 Workplan

The first month will be dedicated to setup, convert and initiate the CG force field testing. During the second month the testing phase will be completed and we proceed to the first docking trials and result evaluation, this stage will be executed iteratively with the scoring function optimization. Optimization will be concluded on the last month and the CG Protein-DNA/RNA docking routine will be implemented followed by the redaction of a report and the research article (Table 1).

Table 1 – Proposed workplan

	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
Convert CG topology into CNS format	**X**					
Convert Protein-DNA/RNA dataset to CG	**X**					
Add and test forcefield in HADDOCK		**X**				
Run docking trials		**X**				
Evaluate results			**X**			
Optimize scoring functions			**X**	**X**		
Implement CG Protein-DNA/RNA docking				**X**	**X**	
Write report and research article					**X**	**X**

5 Methods ans Result Analysis

The overview workflow of the proposed methods is presented in Figure 5.

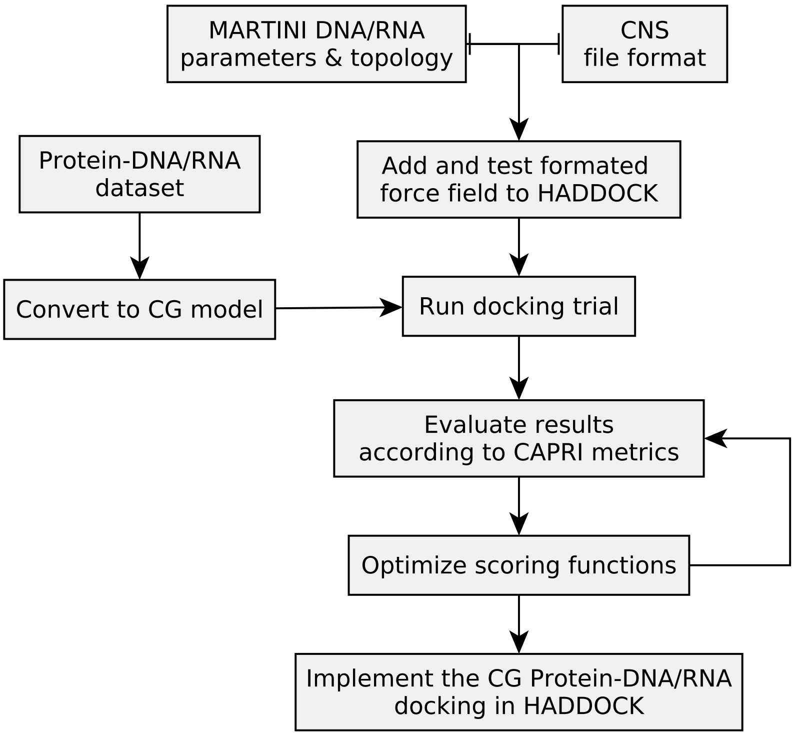

MARTINI DNA/RNA parameters & topology
CNS file format
Protein-DNA/RNA dataset
Add and test formated force field to HADDOCK
Convert to CG model
Run docking trial
Evaluate results according to CAPRI metrics
Optimize scoring functions
Implement the CG Protein-DNA/RNA docking in HADDOCK

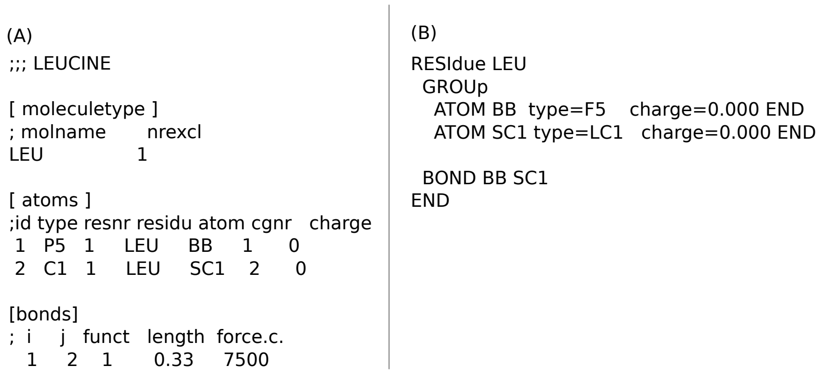
Figure 5 – Overview of the proposed implementation.

5.1 All-atoms to CG conversion

The conversion of the structures to CG models will be carried out by a conversion script adapted from the one provided by the research group responsible for MARTINI. It will be adapted to output distance restraints between the atoms mapped to one CG particle and the other CG particle. These distance restraints are use in an original procedure in HADDOCK to map back the CG model to an all-atom models, while allowing for conformational changes.

5.2 Topology and parameter conversion

The first step of the implementation is to convert the MARTINI DNA/RNA topology and parameters to the CNS format used by HADDOCK. The topology is consisted of information such as, description of atoms, assignment of covalent bonds, bond



```
(A)
;;; LEUCINE

[ moleculetype ]
; molname      nrexcl
LEU            1

[ atoms ]
;id type resnr residu atom cgnr   charge
 1  P5   1     LEU     BB     1      0
 2  C1   1     LEU     SC1    2      0

[bonds]
;  i     j   funct   length  force.c.
   1     2    1       0.33     7500
```

```
(B)
RESIdue LEU
  GROUp
    ATOM BB  type=F5    charge=0.000 END
    ATOM SC1 type=LC1   charge=0.000 END

  BOND BB SC1
END
```

Figure 6 – Example of MARTINI (A) and CNS (B) topology for the amino acid leucine.

angles, charge and others that are combined in order to form the macromolecule. The parameters correspond to force constants and equilibrium values for the various energy terms.

5.3 Docking trials

After integration of the MARTINI CG DNA/RNA force field into HADDOCK, its performance will be evaluated using a benchmark of Protein-DNA complexes (DIJK; BONVIN, 2008). The benchmark (v1.3) is comprised of 47 test cases divided into classes of cases; easy (13), intermediate (22) and difficult (12). Its difficulty is set according to the degree of structural rearrangement after complex formation. Most of DNA binding protein types are covered according to the Luscombe classification (LUSCOMBE et al., 2000).

Docking will be performed using the same information about interfaces as used in the original benchmark. For this, sxperimental or computational data about the interaction interface can be provided to generate *Ambiguous Interaction Restraints* (AIRs) between potential interacting residues (DOMINGUEZ; BOELENS; BONVIN, 2003). These AIRs are responsible for directing the docking process by defining active residues, potentially involved or passive residues, the solvent-accessible neighbours of the active residues. AIRs are defined for all active residues by calculating its effective distances, d^{eff} as follows:

$$d^{eff}_{iAB} = \left(\sum_{M_{iA}=1}^{N_{resB}} \sum_{K=1}^{N_{atoms}} \sum_{N_{kB}=1}^{N_{atoms}} \frac{1}{d^{6}_{m_{iA}\ n\ kB}} \right)^{\frac{-1}{6}} \tag{5.1}$$

Where N_{atoms} represents all atoms of a certain residue in both units of the interacting complex, N_{resB} is the sum of active and passive residues, i is the interaction of all restrains; the $\frac{-1}{6}$ sum reassembles the Lennard-jones potential.

5.4 Result evaluation

The generated docking solutions will be assessed based on CAPRI (JANIN, 2002) criteria, measuring two root mean square deviations (RMSD) calculated over the backbone atoms. Resulting complexes will be categorized into Incorrect, Acceptable, Medium and High according to their interface RMSD (i-RMSD) and ligand RMSD (l-RMSD) (Table 2).

Table 2 – CAPRI categorization of macromoecular complexes

Category	RMSD Å
Incorrect	i-RMSD $\geq$ 4.0 or l-RMSD $\geq$ 10.0
Acceptable	i-RMSD $\leq$ 4.0 or l-RMSD $\leq$ 10.0
Medium	i-RMSD $\leq$ 2.0 or l-RMSD $\leq$ 5.0
High	i-RMSD $\leq$ 1.0 or l-RMSD $\leq$ 1.0

5.5 DNA/RNA scoring function optimization

Each step of the HADDOCK's docking protocol; *Rigid body*, *Flexible* and *Refinement*, has its own scoring function defined as the weighted sum of interaction energies and buried surface area (BSA):

$$\textit{Rigid body (it0)} = 0.01E_{air} + 0.01E_{vdw} + 1.0E_{elec} + 1.0E_{desolv} - 0.01BSA \tag{5.2}$$

$$\textit{Flexible (it1)} = 0.01E_{air} + 1.0E_{vdw} + 1.0E_{elec} + 1.0E_{desolv} - 0.01BSA \tag{5.3}$$

$$\textit{Refinement (water)} = 0.1E_{air} + 1.0E_{vdw} + 0.2E_{elec} + 1.0E_{desolv} \tag{5.4}$$

Optimization of the scoring functions will consist of an evaluation of different ranges of weights in each of interaction energies and comparison with the CAPRI evaluation in order to obtain optimal weights to correctly identify the best generated complexes (Figure 7) and maximize the number of near-native models selected at the rigid-body docking stage. Final optimal weights are based on the average of optimal weight for each conformation for each complex in the benchmark.

Previous studies have shown that protein-DNA binding at the minor groove is driven by the large entropic release of ordered water despite unfavourable enthalpy. Nonetheless, this effect is not representative of a hydrophobic force since water ordering in the DNA is not determined by the apolar groups. This ordering is a consequence of the arrangement of polar groups to stabilize ice-like water organization in the groove (PRIVALOV et al., 2007). Thus, it will be important to also check if desolvation effects can be neglected.

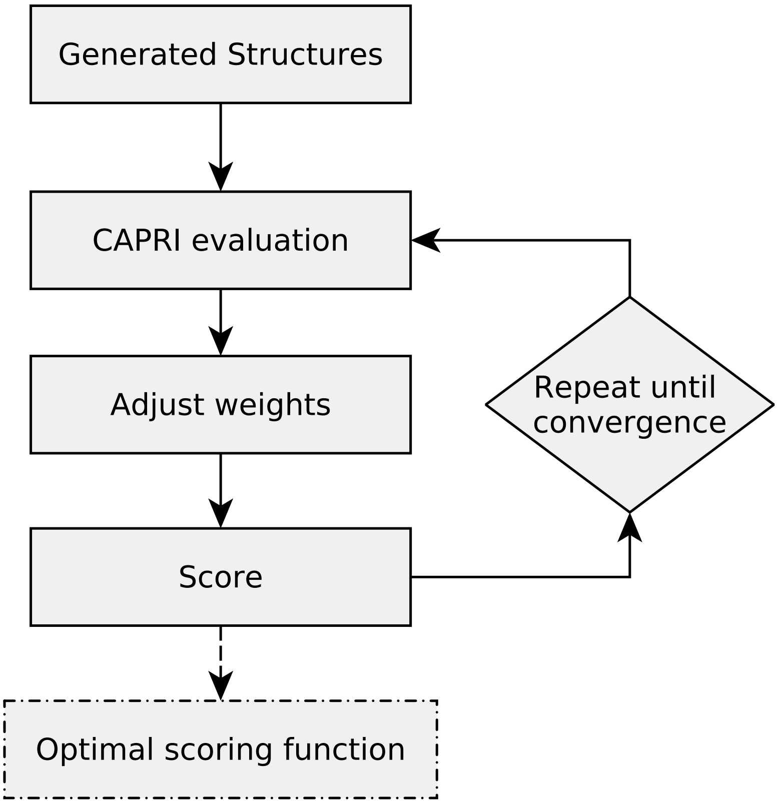


Figure 7 – Diagram of the score optimization

Bibliography

BERMAN, H. M. The protein data bank. *Nucleic Acids Research*, Oxford University Press (OUP), v. 28, n. 1, p. 235–242, jan 2000. Disponível em: <https://doi.org/10.1093%2Fnar%2F28.1.235>. Cited on page: 16

BRUNGER, A. T. CNS (crystallography and NMR system). In: *Encyclopedia of Biophysics*. Springer Nature, 2013. p. 326–327. Disponível em: <https://doi.org/10.1007%2F978-3-642-16712-6_318>. Cited on page: 17

DIJK, M. van; BONVIN, A. M. J. J. A protein-DNA docking benchmark. *Nucleic Acids Research*, Oxford University Press (OUP), v. 36, n. 14, p. e88–e88, jun 2008. Disponível em: <https://doi.org/10.1093%2Fnar%2Fgkn386>. Cited on page: 26

DOMINGUEZ, C.; BOELENS, R.; BONVIN, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, American Chemical Society (ACS), v. 125, n. 7, p. 1731–1737, feb 2003. Disponível em: <https://doi.org/10.1021%2Fja026939x>. Cited on page: 17, 26

HOFFMANN, N. A. et al. Molecular structures of unbound and transcribing RNA polymerase III. *Nature*, Springer Nature, v. 528, n. 7581, p. 231–236, nov 2015. Disponível em: <https://doi.org/10.1038%2Fnature16143>. Cited on page: 19

INGÓLFSSON, H. I. et al. The power of coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Wiley-Blackwell, v. 4, n. 3, p. 225–248, aug 2013. Disponível em: <https://doi.org/10.1002%2Fwcms.1169>. Cited on page: 19

JANIN, J. Welcome to CAPRI: A critical assessment of PRedicted interactions. *Proteins: Structure, Function, and Genetics*, Wiley-Blackwell, v. 47, n. 3, p. 257–257, apr 2002. Disponível em: <https://doi.org/10.1002%2Fprot.10111>. Cited on page: 17, 27

JOHNSON, M. *Numerical computer methods*. Amsterdam Boston: Elsevier Academic Press, 2004. ISBN 9780080497211. Cited on page: 12

JORGENSEN, W. L.; TIRADO-RIVES, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, American Chemical Society (ACS), v. 110, n. 6, p. 1657–1666, mar 1988. Disponível em: <https://doi.org/10.1021%2Fja00214a001>. Cited on page: 17

KMIECIK, S. et al. Coarse-grained protein models and their applications. *Chemical Reviews*, American Chemical Society (ACS), v. 116, n. 14, p. 7898–7936, jul 2016. Disponível em: <https://doi.org/10.1021%2Facs.chemrev.6b00163>. Cited on page: 5, 11, 13

KOLINSKI, A. et al. An efficient monte carlo model of protein chains. modeling the short-range correlations between side group centers of mass. *The Journal of Physical Chemistry B*, American Chemical Society (ACS), v. 102, n. 23, p. 4628–4637, jun 1998. Disponível em: <https://doi.org/10.1021%2Fjp973371j>. Cited on page: 12

KURCINSKI, M. et al. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Research*, Oxford University Press (OUP), v. 43, n. W1, p. W419–W424, may 2015. Disponível em: <https://doi.org/10.1093%2Fnar%2Fgkv456>. Cited on page: 12

LEVITT, M. A simplified representation of protein conformations for rapid simulation of protein folding. *Journal of Molecular Biology*, Elsevier BV, v. 104, n. 1, p. 59–107, jun 1976. Disponível em: <https://doi.org/10.1016%2F0022-2836%2876%2990004-8>. Cited on page: 12

LEVITT, M.; WARSHEL, A. Computer simulation of protein folding. *Nature*, Springer Nature, v. 253, n. 5494, p. 694–698, feb 1975. Disponível em: <https://doi.org/10.1038%2F253694a0>. Cited on page: 12

LINGE, J.; ODONOGHUE, S.; NILGES, M. Automated assignment of ambiguous nuclear overhauser effects with ARIA. In: *Methods in Enzymology*. Elsevier BV, 2001. p. 71–90. Disponível em: <https://doi.org/10.1016%2Fs0076-6879%2801%2939310-2>. Cited on page: 17

LIWO, A. et al. A unified coarse-grained model of biological macromolecules based on mean-field multipole–multipole interactions. *Journal of Molecular Modeling*, Springer Nature, v. 20, n. 8, jul 2014. Disponível em: <https://doi.org/10.1007%2Fs00894-014-2306-5>. Cited on page: 12

LIWO, A. et al. A united-residue force field for off-lattice protein-structure simulations. i. functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *Journal of Computational Chemistry*, Wiley-Blackwell, v. 18, n. 7, p. 849–873, may 1997. Disponível em: <https://doi.org/10.1002%2F%28sici%291096-987x%28199705%2918%3A7%3C849%3A%3Aaid-jcc1%3E3.0.co%3B2-r>. Cited on page: 19

LOUHIVUORI, M. et al. Release of content through mechano-sensitive gates in pressurized liposomes. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 107, n. 46, p. 19856–19860, nov 2010. Disponível em: <https://doi.org/10.1073%2Fpnas.1001316107>. Cited on page: 19

LUSCOMBE, N. M. et al. *Genome Biology*, Springer Nature, v. 1, n. 1, p. reviews001.1, 2000. Disponível em: <https://doi.org/10.1186%2Fgb-2000-1-1-reviews001>. Cited on page: 26

MARK, P.; NILSSON, L. Structure and dynamics of the TIP3p, SPC, and SPC/e water models at 298 k. *The Journal of Physical Chemistry A*, American Chemical Society (ACS), v. 105, n. 43, p. 9954–9960, nov 2001. Disponível em: <https://doi.org/10.1021%2Fjp003020w>. Cited on page: 17

MARRINK, S. J. et al. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, American Chemical Society (ACS), v. 111, n. 27, p. 7812–7824, jul 2007. Disponível em: <https://doi.org/10.1021%2Fjp071097f>. Cited on page: 12, 14, 15

MARRINK, S. J.; TIELEMAN, D. P. Perspective on the martini model. *Chemical Society Reviews*, Royal Society of Chemistry (RSC), v. 42, n. 16, p. 6801, 2013. Disponível em: <https://doi.org/10.1039%2Fc3cs60093a>. Cited on page: 5, 13, 14

PERFORMANCE Of Haddock In Casp Capri. 2015. <http://www.bonvinlab.org/news/Top-Performance-of-HADDOCK-in-CASP-CAPRI/>. Cited on page: 17

PERIOLE, X. et al. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *Journal of the American Chemical Society*, American Chemical Society (ACS), v. 129, n. 33, p. 10126–10132, aug 2007. Disponível em: <https://doi.org/10.1021%2Fja0706246>. Cited on page: 19

PRIVALOV, P. L. et al. What drives proteins into the major or minor grooves of DNA? *Journal of Molecular Biology*, Elsevier BV, v. 365, n. 1, p. 1–9, jan 2007. Disponível em: <https://doi.org/10.1016%2Fj.jmb.2006.09.059>. Cited on page: 27

SAMBRISKI, E. J.; SCHWARTZ, D. C.; PABLO, J. J. de. Uncovering pathways in DNA oligonucleotide hybridization via transition state analysis. *Proceedings of the National Academy of Sciences*, Proceedings of the National Academy of Sciences, v. 106, n. 43, p. 18125–18130, oct 2009. Disponível em: <https://doi.org/10.1073%2Fpnas.0904721106>. Cited on page: 19

SILVA, I. T. G. da; OLIVEIRA, P. S. L. de; SANTOS, G. M. Featuring the nucleosome surface as a therapeutic target. *Trends in Pharmacological Sciences*, Elsevier BV, v. 36, n. 5, p. 263–269, may 2015. Disponível em: <https://doi.org/10.1016%2Fj.tips.2015.02.010>. Cited on page: 19

UUSITALO, J. J. et al. Martini coarse-grained force field: Extension to DNA. *Journal of Chemical Theory and Computation*, American Chemical Society (ACS), v. 11, n. 8, p. 3932–3945, aug 2015. Disponível em: <https://doi.org/10.1021%2Facs.jctc.5b00286>. Cited on page: 5, 15, 16

VENDRUSCOLO, M.; DOBSON, C. M. Protein dynamics: Moores law in molecular biology. *Current Biology*, Elsevier BV, v. 21, n. 2, p. R68–R70, jan 2011. Disponível em: <https://doi.org/10.1016%2Fj.cub.2010.11.062>. Cited on page: 11

WILSON, C.; DONIACH, S. A computer model to dynamically simulate protein folding: Studies with crambin. *Proteins: Structure, Function, and Genetics*, Wiley-Blackwell, v. 6, n. 2, p. 193–209, 1989. Disponível em: <https://doi.org/10.1002%2Fprot.340060208>. Cited on page: 12