Advances in Colloid and Interface Science 232 (2016) 36–48

Contents lists available at ScienceDirect

Advances in Colloid and Interface Science

journal homepage: www.elsevier.com/locate/cis

Historical perspective

Multiscale coarse-grained modelling of chromatin components: DNA and the nucleosome

Nikolay Korolev [a], Lars Nordenskiöld [a,*], Alexander P. Lyubartsev [b,*]

[a] School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551, Singapore
[b] Division of Physical Chemistry, Department of Materials and Environmental Chemistry, Arrhenius Laboratory, Stockholm University, 106 91 Stockholm, Sweden

ARTICLE INFO

Available online 18 February 2016

Keywords:
Macromolecular assembly
DNA condensation
Polyelectrolytes
Computer simulations
Inverse Monte Carlo

ABSTRACT

To model large biomolecular systems, such as cell and organelles an atomistic description is not currently achievable and is not generally practical. Therefore, simplified coarse-grained (CG) modelling becomes a necessity. One of the most important cellular components is chromatin, a large DNA–protein complex where DNA is highly compacted. Recent progress in coarse graining modelling of the major chromatin components, double helical DNA and the nucleosome core particle (NCP) is presented. First, general principles and approaches allowing rigorous bottom-to-top generation of interaction potentials in the CG models are presented. Then, recent CG models of DNA are reviewed and their adequacy is benchmarked against experimental data on the salt dependence of DNA flexibility (persistence length). Furthermore, a few recent CG models of the NCP are described and their application for studying salt-dependent NCP–NCP interaction is discussed. An example of a multiscale approach to CG modelling of chromatin is presented where interactions and self-assembly of thousands of NCPs in solution are observed.

Contents

1. Introduction

This topical review describes applications of coarse-grained (CG) modelling to the description of the two major components of chromatin, DNA and the nucleosome core particle (NCP). Before discussing different coarse-grained models of chromatin components, it is useful to give a brief overview on chromatin structure and its basic physicochemical properties. In most eukaryotes, chromatin is highly uniform at the first level of its organization and can be considered as a linear array of uniform structural units, nucleosomes, formed by DNA and an octamer of highly conserved histone proteins. The histone octamer (HO) comprises two copies each of the four core histone proteins, H2A, H2B, H3 and H4. The nucleosome core particle includes 145–147 bp DNA wrapped as a left-handed 1.75 turn super helix around the histone octamer [1–3]. The NCP is a polyanion–polycation complex with a net charge of about −148e, having a highly negatively charged central particle (−236e, with −294e from DNA and +58e from the globular part of the HO) to which the eight flexible and positively charged N-terminal tails are attached (with net charge +88e). The stability of such a DNA–octamer complex with excess negative charge (also known also as

Abbreviations: CG, coarse-grained; $CoHex^{3+}$, Co(hexammine)$^{3+}$; DH, Debye–Hückel; HO, histone octamer; IBI, iterative Boltzmann inversion; MC, Monte Carlo; IMC, inverse Monte Carlo; MD, molecular dynamics; NCP, nucleosome core particle; oxDNA, "Oxford" CG model of DNA; RDF, radial distribution function; SCG, super-coarse-grained; SPN, sugar, phosphate and nuclear base.

* Corresponding authors.
E-mail addresses: larsnor@ntu.edu.sg (L. Nordenskiöld), alexander.luybartsev@mmk.su.se (A.P. Lyubartsev).

http://dx.doi.org/10.1016/j.cis.2016.02.002
0001-8686/

overcharging effect) at physiological salt conditions can be understood by electrostatic theories [4–7].

On the next level of chromatin organization linker DNA of a variable length (10–70 bp) connects the NCPs to form an array of chromatin. In vitro, the nucleosome arrays are folded yielding a fibre of approximately 30 nm diameter whose detailed structure is still a matter of debate [8–14]. The relevance of the 30-nm fibre in vivo has also recently been the subject of discussions [15–17] (and references cited in [17]). Inter-array aggregation leads to additional higher-order compaction, resulting in chromosomes. In higher organisms, one more histone protein, the linker histone H1, is present at a NCP:H1 ratio of 1:1 [18]. The linker histones contribute to the chromatin structure stabilization but it was shown that the nucleosome arrays can adopt the folded compact state without histone H1 [18,19].

Eight flexible and basic N-terminal histone tails protrude out from the core domain of the NCP. These facilitate interactions between neighbouring nucleosomes [20–31]. Chromatin is like DNA, a highly negatively charged polyelectrolyte. Observations in vitro show that folding of the array and further inter-array aggregation (often called oligomerization or self-association) into tertiary chromatin structures can occur as a result of increased monovalent salt concentration or by the addition of Mg^{2+} or other multivalent cations [21,23,26,28–35]. This illustrates the dominance of electrostatic interactions in the compaction of chromatin [36–38]. NCPs in solution display a similar behaviour [20,22,25,39–41]. Systems of tailless (trypsin treated) NCPs do not show Mg^{2+}-induced folding/aggregation, which shows the important role that the N-terminal tails execute in condensation of NCPs [34]. The knowledge about the mechanisms of chromatin unfolding, stretching, DNA unwrapping from the histone core has advanced in recent years [42–44], but the physical mechanism that determines these processes still remains to be fully understood.

A common feature of condensed nucleosomes and folded compact chromatin structures is the presence of nucleosome stacking, i.e. close NCP–NCP contacts between the flat surfaces of the histone octamer core on both sides of the cylindrical wedge-shaped NCP. This has been experimentally observed in NCP crystals [1,3,45–47], in NCP liquid crystalline phases [41,48–51], in the crystal of the tetranucleosome [9], in folded nucleosome arrays [8,10,52], and in cryo-microscopy images of frozen isolated native chromatin [53,54].

Chromatin structure depends critically not only on the ionic conditions modulating the electrostatic interactions, but also on other factors such as nucleosome repeat length (NRL), presence of linker histone, histone variants and nature and degree of histone tail post-translational modifications. To investigate all these factors by coarse-grained computer modelling is particularly challenging as it requires rigorous force fields that capture electrostatic, hydrophobic, specific (hydrogen bonding) as well as solvent interactions.

In recent decades computer simulations have become a well-established tool for investigation of molecular structure. Molecular simulations based on Monte Carlo (MC) or Molecular Dynamics (MD) technique can provide detailed, atomistic resolution, structural and dynamical information about the system in strict agreement with the fundamentals of the statistical mechanics. Molecular simulations are especially important in studies of soft matter systems, where the absence of regular structure and temporal fluctuations makes it difficult, if not impossible, to access such information by any experimental technique. Information obtained in molecular simulations thus becomes complementary to experiments, and provides an important link between experimental observations and theoretical description of the studied systems.

In many cases of biomolecular systems, such as DNA–nucleosome complexes, an atomistic description is not practical as the number of particles would be very large. Simplified coarse-grained models become necessary, allowing a coarser description of the molecules and interactions when moving towards description at larger length and time scales. In a typical CG model, atoms of the macromolecules are united into coarse-grained sites and solvent atoms are often not considered explicitly. This reduces greatly the number of degrees of freedom of the studied system and allows simulations of much larger systems which are not feasible to simulate at the atomistic level. Studies of models which can be characterized as "coarse-grained" began already at the earlier stages of molecular modelling fifty years ago (when the term "coarse-grained" was not used at all). For example, the primitive electrolyte model was used to represent hydrated ions as charged spheres in a dielectric medium with a dielectric permittivity corresponding to water [55,56], and a simple freely-jointed chain model of a polymer [57] was used to model polymers in solution. Later this class of models resulted in the appearance of various continuum solvent polyelectrolyte models describing e.g. distribution and competitions of ions around DNA [58–63]. In application to chromatin such simple models have also been used, see the review [64].

A basic problem in coarse-grain modelling of molecular systems is the choice of interaction potential between coarse-grained units of macromolecules. In the so-called *top-down* methodology one would start from a simple model defined by physical principles, e.g. introducing a Coulombic interaction potential between the ions scaled by the water dielectric permittivity and fitting parameters to reproduce some experimental properties (for example setting the radii of ions to reproduce experimental osmotic coefficients data). Within the *bottom-up* multi-scale scheme, effective coarse-grained potentials are derived from atomistic simulations (and in some cases from ab-initio computations). The atomistic force fields reflect the real chemical structure of the studied system and they are generally more established than coarse-grained force fields. The interactions between the CG sites are represented by effective potentials that include the effects of the water solvent and which are deduced from the atomistic simulations. This enables the modelling of very large molecular assemblies with a reduction of the number of particles by several orders of magnitudes, but still using effective forces that implicitly include a detailed description at the atomic level. Simulations of such large systems can then be performed over a range of time scales therefore being a multi-scale approach.

We describe the present status of multi-scale CG modelling of DNA and the NCP and the relation of these models to experimental studies. We review coarse grained modelling of the major component of chromatin, DNA; then we describe the results of CG modelling of the nucleosome core particle (NCP). Approaches and models that have been used in recent years are discussed. A main focus will be on bottom-up multiscale computer simulation methods based on parameterization of coarse-grained potentials from atomistic trajectories that describe the ensemble of states of the system from which various average structural properties can be calculated. There are several approaches to such systematic CG modelling, namely structure based approaches like the inverse Monte Carlo method [65–67] and the iterative Boltzmann inversion [68], the force matching method [69,70] and the relative entropy minimization methods [71]. Another group of CG approaches are based on principles similar to ones used in the construction of the atomistic force fields, by parameterization of interaction potentials using experimental data like thermodynamic properties, e.g. the MARTINI CG force field approach [72].

2. Theoretical foundation for bottom-up multiscale coarse-grained modelling

Generally, the problem of going from a high-resolution (atomistic) description to a coarse-grained one can be formulated as follows. The atomistic model is defined in term of a specific chemical structure of the system, including the atomic composition, chemical bonds and an atomistic force field defining the interactions between the atoms. We assume that we are able to model such a system by atomistic MD and obtain a statistically significant trajectory, which is a set of system configurations which can be used to extract ensemble averages of various physical properties. A coarse-graining procedure can be introduced,