European Science Foundation
SimBioMa Exchange Grant
September 2010

Scientific Report

Novel coarse-grained force field for DNA-protein recognition

Grantee:
Aymeric Naômé University of Namur (BE)
(supervisor: Daniel P. Vercauteren, Prof.)

Host person:
Aatto Laaksonen, Prof. Stockholm University (SE)

1 Purpose of the visit

Computer simulation is the great tool to investigate complex biomolecular systems whose structure and dynamics at the molecular level are sorely accessible to conventional experimental techniques. Unfortunately, in the case of systems composed of DNA, solvent, ions, and proteins, the phenomenologically relevant size and time scale are in the order of dozens of nanometres and up to the second. Modelling such systems at the atomic level of detail is far too demanding in terms of computer and time resources. The development of coarse-grained (CG) models of biological systems for efficient and accurate large scale molecular simulations has thus recently become a subject of extensive interest in biophysical sciences. The understanding of the formation mechanism, structure, and function of protein-DNA biological assemblies in particular can benefit a lot from CG simulations. However, even though a wide range of CG models is already available for most biological molecules (a review on models and methods is available as ref. 1), only a few implementations are reported concerning DNA [2-7] which have moreover not reached yet a satisfactory level of reliability for performing useful (in our meaning) molecular dynamics (MD) studies together with proteins. For this purpose, we need to develop CG representations with the following minimal characteristics: (i) the nucleobase and amino acid sequence information has to be retained as much as one can to render the protein-DNA association specificity, (ii) peptide and nucleic acid backbones must be flexible, and (iii) the solvent and ions are modelled explicitly. To begin, we selected two different approaches to create a CG force field (FF) for DNA that satisfies those features: the MARTINI model [8-11] and the inverse Monte-Carlo (IMC) method [12-20].

The IMC method has been developed by Prof. A. P. Lyubartsev and Prof. A. Laaksonen at Stockholm University about fifteen years ago [13] and provides an elegant iterative way of solving the inverse problem of statistical mechanics, viz. reconstructing the interaction potentials from a distribution of matter, here a set of radial distribution functions (RDFs). The theory is based on a theorem due to Henderson [12] stating the uniqueness of the potential inverted from a given RDF under given conditions of temperature and density. The applicability of the theory was illustrated with the parametrisation of effective potentials for water molecules and water-ion interactions from *ab initio* MD [16,17], solvent mediated ion-ion [14,18] and ion-DNA [15], one-site model of water, L- and D- proline in DMSO [20], and a CG lipid model [19], all from classical MD. Those applications have proven successful but the capability of the method to yield potentials suited for performing simulations with large complex biomolecules has not been fully tested yet.

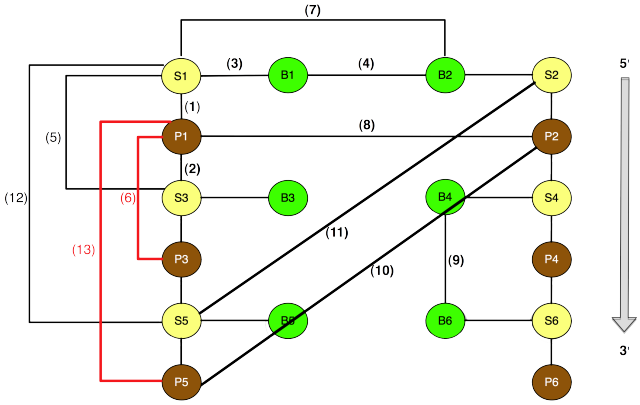
The aim of our visit to Stockholm was thus to construct reduced CG representations of DNA, taking advantage of the expertise of the method's originators, the IMC codes developed and maintained on-site by Mr. A. Mirzoev, and microseconds of high quality B-DNA MD trajectories at

our disposal (kindly given by Prof. Lavery and collaborators).

Those trajectories consists in a set of fifty-nanosecond MD simulations of 39 octadecamer B-DNA fragments of different sequence containing the 136 unique tetranucleotide combinations [21]. This raw material provides us with enough conformational sampling to confidently construct sequence-specific CG models taking into account up to the first-neighbour context of a base pair (i.e., triplets).

2 Description of the work carried out during the visit

The first step of the project work was to build a test model of CG DNA based on a three-site topology regardless of the nucleobase sequence. The centre of mass of the phosphate group (P), sugar moiety (S), and nucleobase fragment (B) were assigned to a CG interaction site. An additional site was introduced for the counterion (Na). Seventeen effective pairwise potentials link the CG sites together. The topology and interactions are sketched in Figure 1. One hundred and fifty nanoseconds of all-atom (AA) MD trajectories from a precedent work [22] were mapped with this topology. The RDFs corresponding to the seventeen pairwise interactions were calculated using the `rdf` utility of the MD program suite MDynaMix [23]. The collected RDFs are input of the IMC module itself (dubbed `MOLSIM`, written by A. Mirzoev and still under development). The iterative procedure proceeds as follows. The discretised potentials of mean force are calculated as the trial potentials, a Monte-Carlo (MC) simulation is then performed, the RDFs are re-calculated and the starting potential is corrected as to minimise the difference between the original and the MC RDFs. The whole procedure resembles the Newton inversion solution of a nonlinear multidimensional equation (the detailed procedure and statistical mechanics relationships are found in ref. 13). The iterative Newton inversion is preceded by several stages of iterative Boltzmann inversion as described by A. K. Soper [24], a method that converges faster for large discrepancies of the potentials. The purpose of this test system was to determine the optimal set of computational parameters to achieve the best and fastest convergence of the RDFs.



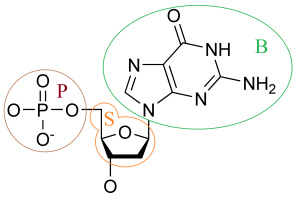


Figure 1. Topology of the three-site thirteen-interaction test model. The thirteen pairs are labelled and named as follows. (1) S-P, (2) P-S, (3) S-B, (4) B-B pairing, (5) S-S strand, (6) P-P strand, (7) S-B opposite, (8) P-P opposite, (9) B-B strand, (10) P-P twist, (11) S-S twist, (12) S-S rigidity, and (13) P-P rigidity. The intermolecular potentials not shown are B-Na, S-Na, P-Na, and Na-Na.

In parallel, we started the pre-processing of the B-DNA MD trajectories received from Prof. R. Lavery in the course of the month of June 2010. We adapted, implemented, and created tools for manipulating this huge body of data and made it suitable for the IMC procedure. We mapped the two microseconds of trajectory following a crude one-site-per-nucleotide scheme with and without nucleotide identity distinction and also initiated the IMC procedure.

We have also devoted a part of the research time to a second approach aforementioned. The MARTINI model is a now well established procedure for deriving CG models of various molecules of biological interest. For example, lipids [9], polypeptides [10], and sugars [11] have already been successfully modelled. In a few words, the method consists in mapping molecules with a limited set

of building blocks so as to best fit thermodynamic values such as the partition coefficients between an organic phase and water. Bonded terms between CGs are then extracted from AA MD simulations of simplified related systems. Applying the MARTINI procedure to nucleic acids summarizes up to three main stages: (i) Map the nucleobases and nucleobase analogues with MARTINI CGs best representing the molecular volume and the different character of polarity of different parts of the molecule. Perform CG MD simulations in cyclohexane/water and chloroform/water to determine partition coefficients. Compare to the available experimental data and secondarily to results of thermodynamic integration (TI) free energy calculations at the AA level. If necessary, refine the CG representation and re-do the calculations until good agreement is reached. (ii) From CG-mapped AA MD of oligonucleotides, extract the bonded terms parameters by inverting the distributions of bonds, bending angles, and torsion angles. (iii) As to retain a well-behaved double-helix, possibly set extra *ad hoc* "backbone twist", "Watson-Crick pairing", and "base-pair step stacking" terms. The persistence length and the sequence/ionic strength dependence of DNA melting temperatures are available target properties. Taking advantage of the existing computing infrastructure at the iSCF (www.iscf.be) and of extensive computer time newly available to us from HPC2N (www.hpc2n.umu.se) and PDC (www.pdc.kth.se), we carried out TI calculations of the free energy of solvation of nucleobase analogues in water, cyclohexane, and chloroform at the AA level for comparison with the CG simulations and experimental data. We also started MD simulations of the 64 different trinucleotides in water as to extract bonded parameters for the MARTINI FF.

3 Description of the main results

The success of the whole IMC procedure depends on different parameters which we optimised for DNA systems.

Regularisation factor. The best convergence of the RDFs is achieved with a regularisation factor of $\delta=0.1$[1]. When convergence is reached with $\delta=0.2$, the RDFs indeed oscillate back and forth around the reference values. This behaviour is encountered specially when the RDFs have long tails and head with values close to zero. The use of trimmed RDFs (below 0.001-0.003) as reference, ignoring low probability states, is thus recommended. Later refinement stages need a low $\delta=0.05$.
Monte-Carlo simulations. The number of MC steps necessary to calculate smooth RDFs depends of course on the number of pair interactions per RDF in the system. In all cases considered so far, 15 to $20\cdot10^6$ steps (including $5\cdot10^6$ for equilibration) are sufficient. A maximum MC displacement of dr=0.5-1.0 Å works well.
Convergence of the iterative procedures. The RDFs determining the convergence of the whole procedure are the intermolecular ones, and the more numerous they are the longer it takes. In general, about 15 inverse Boltzmann iterations (IBI) are required for the deviation to the reference RDF to reach a plateau (provided the number of MC steps and regularisation factor are carefully chosen). Any attempt to move to the inverse Newton iterations (INI) stage without sufficient convergence is bound to fail, the RDFs diverge rapidly eventually and end into noise.

The RDFs extracted from the 150 ns MD simulation of the dodecamer 5'-d(CGCGAGTTCGCG)-3' mapped with the CG topology of Figure 1 are displayed below in Figure 2 together with the converging RDFs calculated from the IMC procedure and the deviation to the reference RDFs. We have first trimmed the reference RDFs, with special attention paid to intramolecular RDFs. Starting from those RDFs, we performed 10 IBI with $20\cdot10^6$ MC steps with $5\cdot10^6$ equilibration steps and regularisation factor $\delta=0.2$ Then, 10 INI with $20\cdot10^6$ MC steps with $5\cdot10^6$ equilibration steps and regularisation factor $\delta=0.1$. were carried out. Finally, 10 additional INI

1 This is the factor scaling the correction brought to the potential: $V_i = V_{i-1} + \delta \Delta V$

with $50 \cdot 10^{-6}$ MC steps with $10 \cdot 10^{-6}$ equilibration steps and regularisation factor δ=0.05 yielded well converged RDFs and potentials appropriate for CG MD simulations. The intermolecular RDFs are still not completely converged but show the same trend as the reference RDFs.

We are now working at preparing the converged potentials for use with test MD simulation of 100's base pairs long DNA fragments and jointly continuing the IMC calculation in order to achieve maximum convergence of the RDFs/potentials. Practically, the potentials have to be interpolated. And the intermolecular potentials need to be damped so that their value and that of the force at the cut-off distance are zero. Since the intermolecular potentials are calculated up to 25 Å, we also need to extend the look-up tables for CG simulations at a larger size scale. This is achieved by adding a *1/εr* Coulombic term up to the cut-off in force.

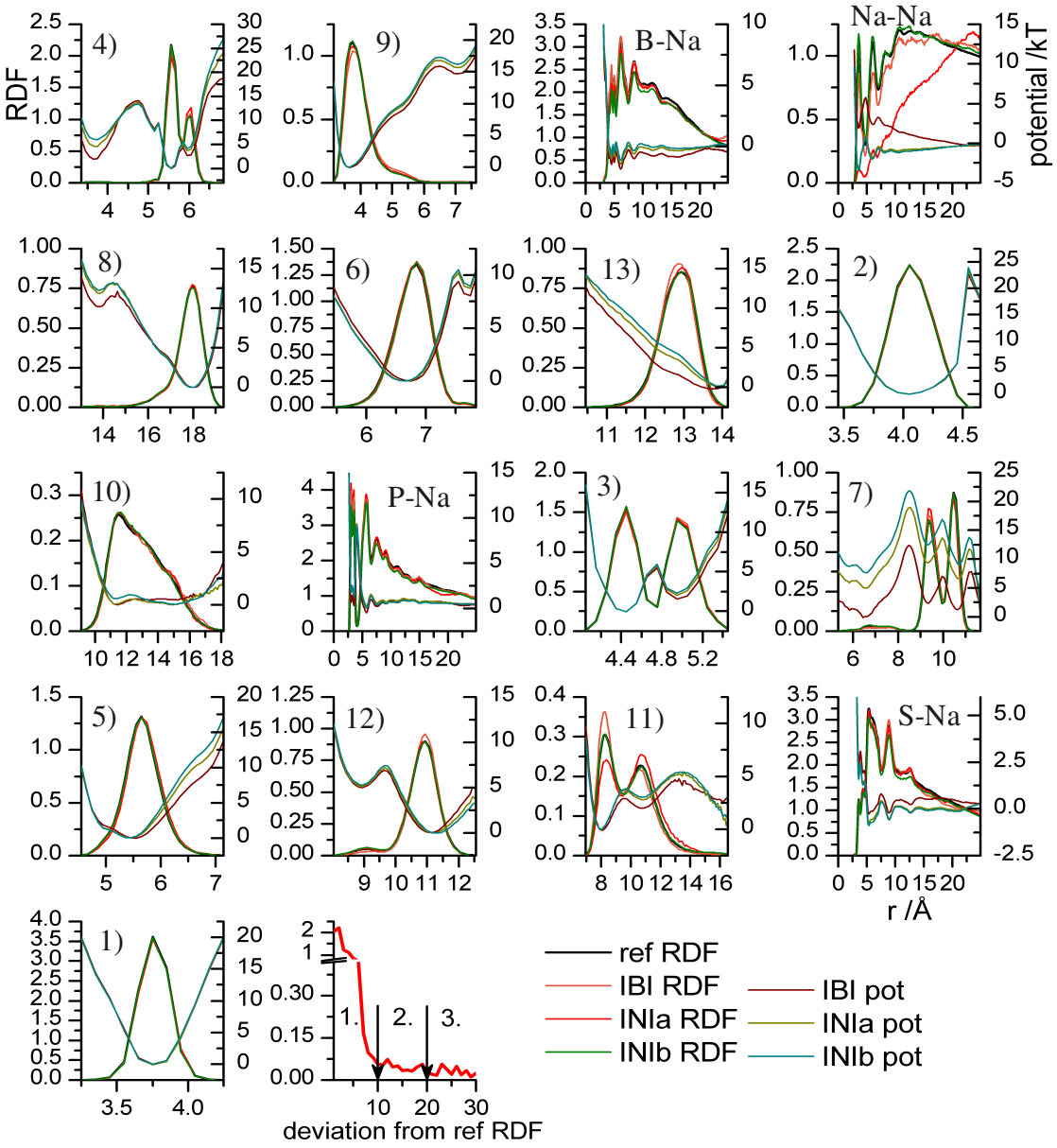


Figure 2. RDFs and potential for the 17 CG pair interactions considered in Fig. 1 plus the deviation to the reference RDFs displayed as a measure of the convergence of the IMC procedure. The 17 first graphs sketch the reference RDFs and the RDFs and potentials as obtained from the last iterations of the three-stage IMC procedure described in the text. The numbering corresponds to that of Fig. 1. The last graph in the bottom pictures the iterative refining of the initial potentials. Zone 1 stands for the 10 initial IBI, zone 2 for the first INI, and zone 2 for the 10 last INI. Units for the potential energy are k_BT per unit interaction.

In the meantime, we considered the octadecamer trajectories from R. Lavery and reduced them to a one-grain-per-nucleotide representation. We calculated the RDFs averaged over the 39 trajectories disregarding the sequence (trimmed below 0.003, with a resolution of 0.05 Å and a cut-off set to 32 Å) and tested several interaction networks. A first model with 7 intramolecular interactions failed at retaining the extended conformation of DNA. Ten interactions were still insufficient. This is because a one-grain model needs more constraints in the form of interaction potentials to depart from the behaviour of a random coil whereas some orientational order is already present in the former three-site test model. The final conformation of the two ill-behaved models after the last MC simulation of a run of 10 IBI following the same specifications as above along with that of a 20-potential model are presented in Figure 3. The iterative procedure is still under-way. Because of the longer range intramolecular interaction considered compared to the previous models (up to 30 Å), we now have 16,000 discrete values of potential to converge consistently, which is equivalent to solving a system of an equal number of non-linear equations (4,250 for the three-site model) ...what is rather time consuming. The RDFs shown in Figure 3 are the reference ones.

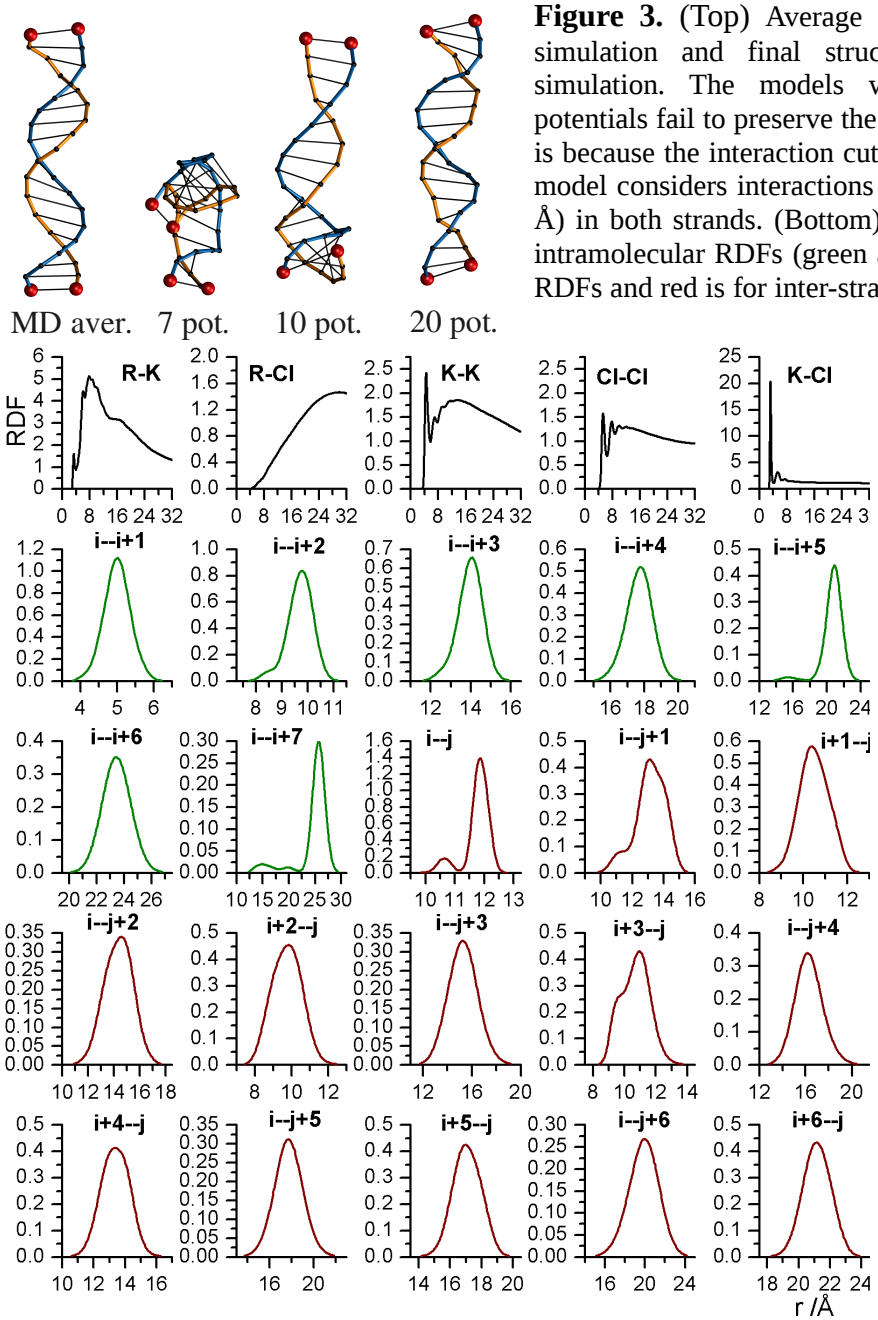
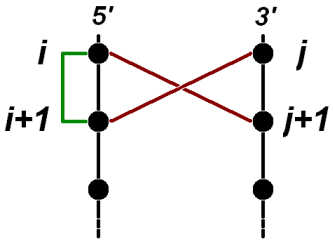


Figure 3. (Top) Average structure from the 2 µs of MD simulation and final structures from the 10th IBI MC simulation. The models with 7 and 10 intramolecular potentials fail to preserve the double helix shape of DNA. This is because the interaction cut-off is too short. The 20-potential model considers interactions as far as 6-7 CG sites apart (~30 Å) in both strands. (Bottom) 5 intermolecular (black) and 20 intramolecular RDFs (green and red); green is for intra-strand RDFs and red is for inter-strand RDFs. The indices assigned to RDFs are explained graphically below. The helicity is taken into account by differentiating 5' and 3' directions.



Free energy of transfer of nucleobase analogues

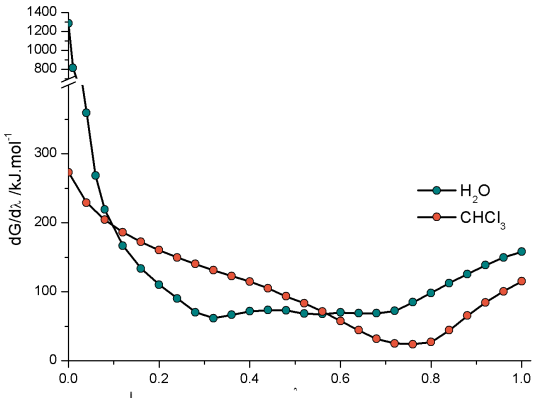
The nucleobases and their *N*-alkylated derivatives being barely soluble in water, it is difficult to determine experimentally their solubilities and free energy of hydration. A more accessible value resides in the partition coefficient from an organic solvent to water, their determination being though strenuous as conveyed by the only two occurrences of experimentally determined values for nucleobases analogues available (to our knowledge) in the literature [25,26]. We have thus calculated the free energy of solvation in water cyclohexane, and chloroform for three nucleobase derivatives, *N*-methylated (*N*-MeX), *N*-butylated (N-BuX), and *N*-tetrahydrofurylated (*N*-ThfX), using the TI approach (Fig. 4). The FF chosen to represent the solute is GROMOS53A6 [27], especially meant to reproduce free energies of solvation, following the same philosophy as the MARTINI FF. The SPC model is used for water [28] and the parameters for chloroform are those from Dietz and Heinzinger [29]. All simulation boxes are 4.5 nm in size with cubic PBC consisting in a solute molecule and a number of solvent molecules reproducing the pure liquid density at the room temperature. *In vacuo* simulations were performed as follows: (i) conjugated gradients energy minimisation with steepest descent every one-hundred steps, (ii) 200 ps NVT equilibration, (iii) 3 ns NVT production run with energy terms stored every ps (1 per 1,000). Simulations in solution follow the sequence: (i) conjugated gradients energy minimisation with steepest descent every one-hundred steps, (ii) 500 ps NVT equilibration, (iii), 500 ps NPT equilibration, (iv) 3 ns production run with energy terms stored every ps. Compressibility coefficient employed are $4.5 \cdot 10^{-5}$, $11.2 \cdot 10^{-5}$, and $9.8 \cdot 10^{-5}$ bar^{-1} for water, cyclohexane, and chloroform, respectively. For cyclohexane and chloroform, like *in vacuo*, a total of 26 MD simulations with switching parameter increment of $\Delta\lambda=0.04$ were carried out. For water, because of the steep slope of the free energy derivative curve for small values of λ, four additional windows were necessary to obtain a smooth curve: λ=0.01; 0.02; 0.03; 0.06 (Fig. 5). All simulations were done with GROMACS [30].

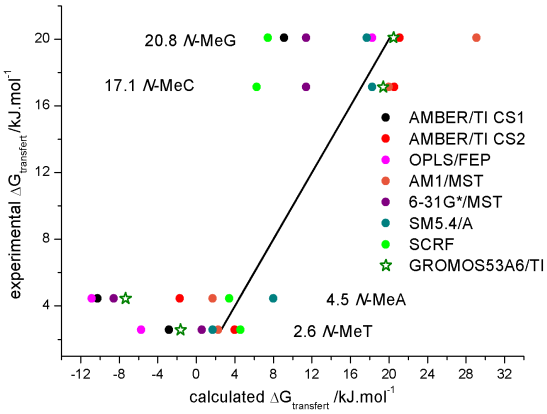
$$\begin{array}{ccc} [dummies]_{vac} & \xrightarrow{\Delta G=0} & [dummies]_{sol} \\ \uparrow{\scriptstyle \Delta G^{1}_{annihilation}} & & \uparrow{\scriptstyle \Delta G^{2}_{annihilation}} \\ [solute]_{vac} & \xrightarrow[\Delta G_{solvation}]{} & [solute]_{sol} \end{array} \qquad \Delta G_{\text{solvation}} = \Delta G^{1}_{annihilation} - \Delta G^{2}_{annihilation}$$

$$\Delta G_{annihilation} = G_{dum.} - G_{sol.} = \int_{\lambda_{sol.}}^{\lambda_{dum.}} \mathrm{d}\lambda \left\langle \frac{\partial \mathrm{U}(\lambda)}{\partial \lambda} \right\rangle_{\lambda} \; ; \; \lambda_{sol.} = 1 \text{ et } \lambda_{dum.} = 0$$

Figure 4. Sketch of the thermodynamic cycle for thermodynamic integration (TI) calculation of the free energy of solvation. It is the sum of two non-physical contributions $\Delta G_{annihilation}$ obtained by gradually switching the interactions of the solute with the solvent to zero with the parameter λ.

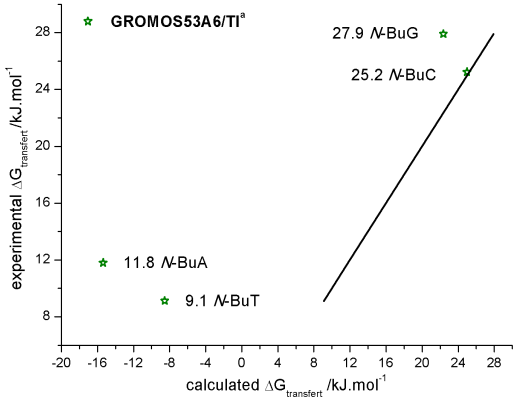
Our results are summarized in the tables and graphs of Figures 5 and 6 together with experimental and other theoretical values [31-35]. Alike many various methods, the free energies of transfer water/chloroform for *N*-methylcytosine (*N*-MeC) and *N*-methylguanine (*N*-MeG) are in good agreement with experiment [25] (Fig. 5). The two other values show a severe discrepancy and even change sign but are to be compared with mediocre results from other methods, especially those based on molecular mechanics. We obtain an averaged unsigned error ranked fourth out of eight. Free energies of hydration are for their part in qualitative agreement with the only experimental values determine for *N*-methyladenine (*N*-MeA) and *N*-methylthymine (*N*-MeT): -56.9 and -(38--53) $kJ.mol^{-1}$ [25] versus -31.9 and -38.7 kJ.mol-1. The very negative values calculated for *N*-MeC and *N*-MeG are compatible with the impossibility of determining those values experimentally due to their very low vapour pressure [25]. The free energies of transfer water-to-cyclohexane for *N*-BuX follow the same trend (Fig. 6). The systematically too negative free energies for A and T derivatives could be due to the overestimated hydrophobicity of adenine and thymine in the GROMOS53A6 FF. Analysis for *N*-ThfX is still in progress.





$\Delta G_{transfert}(w \rightarrow chlo)$ (kJ.mol^{-1})	*N*-MeA	*N*-MeC	*N*-MeG	*N*-MeT	average unsigned error
AMBER/TI CS1[a]	-10.3	11.4	9.1	-2.9	9.2
AMBER/TI CS2[a]	-1.7	20.5	21.1	4.0	3.0
OPLS/FEP[b]	-10.8	20.0	18.3	-5.7	7.1
AM1/MST[b]	1.7	20.0	29.1	2.3	3.7
631G*/MST[b]	-8.6	11.4	11.4	0.6	7.3
SM5.4/A[c]	8.0	18.3	17.7	1.7	2.0
SCRF[d]	3.4	6.3	7.4	4.6	6.6
GROMOS53A6/TI	**-7.4**	**19.4**	**20.5**	**-1.6**	**4.7**
Experiment[e]	4.5	17.1	20.1	2.6	

Figure 5. (Top left) Example of a curve of the derivative of the free energy with respect to the switching parameter λ for *N*-methylguanine (*N*-MeG) in water (blue) and in chloroform (orange). The difference between the areas described by the two curves gives the free energy of transfer from water to chloroform $\Delta G_{transfer}(w \rightarrow chlo)$. Every point corresponds to the derivative of the free energy averaged on 3 ns (*top left*). (Top right and table) The $\Delta G_{transfer}(w \rightarrow chlo)$ calculated in this work for the four *N*-methylated nucleobases (*N*-MeX) gathered as star symbols in the graph and as bold values in the table together with values obtained with other theoretical methods ([a]ref. 35, [b]ref. 33, [c]ref. 34, [d]refs. 31 and 32, [e]ref. 25). The line depicts an ideal match with experimental values.



$\Delta G_{transfert}(w \rightarrow chx)$ (kJ.mol^{-1})	*N*-MeA	*N*-MeC	*N*-MeG	*N*-MeT	average unsigned error
GROMOS53A6/TI	-3.5	36.4	34.5	3.2	--
experiment	--	--	--	--	
	N-BuA	*N*-BuC	*N*-BuG	*N*-BuT	
GROMOS53A6/TI	-15.4	25.0	22.4	-8.6	12.7
experiment[a]	11.8	25.2	27.9	9.1	
	N-ThfA	*N*-ThfC	*N*-ThfG	*N*-ThfT	
GROMOS53A6/TI	tbd	tbd	tbd	tbd	
experiment[a]	17.8	30.4	31.3	17.1	

Figure 6. Free energy of transfer from water to cyclohexane $\Delta G_{transfer}(w \rightarrow chx)$ for the *N*-methyl- (*N*-MeX) and *N*-butylnucleobases (*N*-BuX) ([a]ref. 26). The free energies for the *N*-tetrahydrofurylated (*N*-ThfX) are still under analysis.

MARTINI topologies of the nucleotides

We have assigned the type of MARTINI CG sites representing the four natural nucleotides following chemical intuition. The *4-to-1* rule (i.e., four heavy atoms fused into one super-atom) is not obeyed but is always higher than the *2-to-1* scheme recommended for cyclic planar compounds (G: 2.4/1; A: 2.6/1; T: 2.9/1; C: 2.7/1). The average mass per grain is G: 36.4, A: 39.0, T: 43.3, and

C: 41.1 a.m.u.. Despite the fact that the MARTINI procedure assigns a mass of 72 a.m.u. for all grains (equivalent to the mass of four water molecules) for algorithmic efficacy, it will surely turn necessary to give every CG its nominal mass. The CG are positioned at the centre of mass of the atoms it represents, yet with some flexibility to best fit the molecular volume of the AA model. The trial MARTINI topologies are pictured in Figure 7.

In order to determine the bonded interactions parameters between CG sites, we performed MD simulations of the 64 trinucleotides in a 150 mM ionic strength (KCl) solution. The revised Amber FF Parmbsc0 [36] was employed with the SPC water model together with Dang *et al.* parameters for ions [37]. We acquired 12 ns of trajectory for every trinucleotide system. We have started extracting the force constants and equilibrium values for the simple harmonic bond stretching and angle bending potentials of the MARTINI FF by a simple Boltzmann inversion of the distribution of bonds and valence angles between the CG sites mapped onto the AA structures. In a first time, we did not take the sequence context into account. This stage is still on going but we can already illustrate the procedure for the angle formed by the CG sites Qa-C2/C3-N0 of the sugar-phosphate backbone (Fig. 8). As the population of a given state with a particular angle value is related to the energy difference with the most probable state by its Boltzmann factor, it is possible, by fitting the distribution with a Gaussian curve, to obtain the force constant of the corresponding harmonic potential. The simple expression in the equation of Figure 8 gives the equilibrium angle Φ_0=88.3° and force constant K_Φ=103 J.mol^{-1}.degree^2.

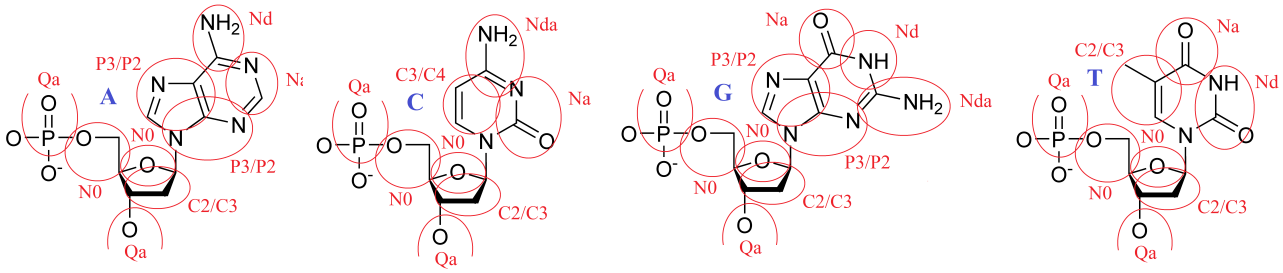
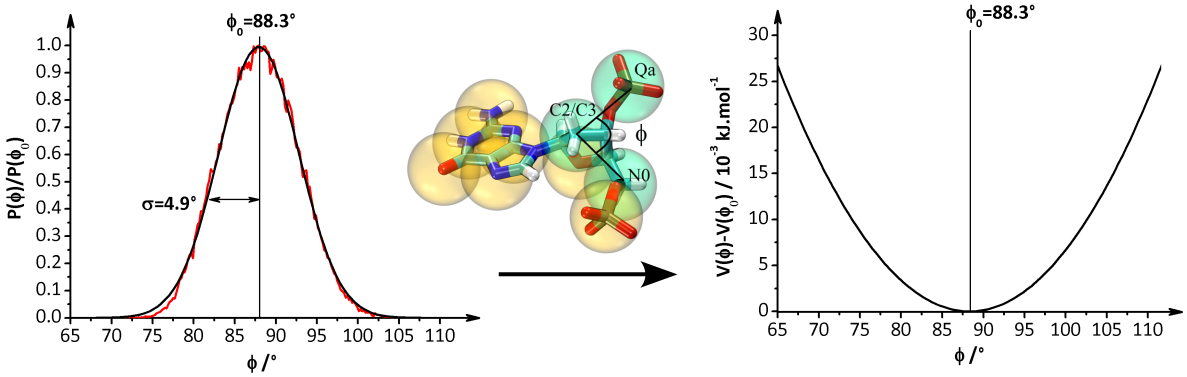


Figure 7. Sketch of the MARTINI topologies schematics for the four nucleotides. The optimal type for the CG sites represented here with two designations will be set after calculations of free energies of transfer.



$$P(\phi)/P(\phi_0)=e^{-V(\phi)-V(\phi_0)/k_BT}=\frac{1}{\sigma\sqrt{2\pi}}e^{-(\phi-\phi_0)/2\sigma^2}$$

$$V(\phi)-V(\phi_0)=-k_BT\ln[P(\phi)/P(\phi_0)]$$

$$=\frac{1}{2}K_\phi(\phi-\phi_0)^2$$

$$\rightarrow K_\phi=\frac{k_BT}{\sigma^2}$$

Figure 8. Illustration of the inversion process for the bonded parmeters. The Φ angle distribution from the MD simulation is fitted with a Gaussian function and then inverted to obtain the corresponding harmonic potential.

4 Meeting and conferences

Part of the work achieved at Stockholm University was presented in a poster entitled "*MARTINI and Inverse Monte Carlo: two capable approaches towards a DNA coarse-grained model retaining the nucleobase sequence information*" on the occasion of the "12ème Réunion des Chimistes Théoriciens Francophones" held in Namur July 4-8 (URL: www.fundp.ac.be/rctf-2010). A Ph.D. seminar was also given to the scientific staff of the physical and inorganic chemistry Divisions.

5 Future collaboration with host institution

The already well-established collaboration between the two groups at Stockholm University and the University of Namur is being reinforced this year with another shared Ph.D. Student from Namur, Mr. Mathieu Fossépré, under the supervision of both group leaders. A long term research stay at Stockholm University is scheduled early 2011. On the IMC hand, the forthcoming work will consist in adjusting the level of reduction of the DNA CG model so that it meets the phenomenological dimensions of the biological processes we aim at studying and thus becomes a practical model. Later on, we also envision to create a library of DNA CG fragments with different level of "coarseness" that could eventually make up a class of multi-purpose DNA CG FF adapted to parallel multi-scale simulations. On the technical side, it is also planned to continue developing modules to give a more user-friendly character to the procedure and ultimately issue an integrated package to coarse-grain trajectories with user-defined topologies and interface with the MD program GROMACS. The `MOLSIM` program will also be implemented for execution in parallel and for adaptive scaling of the regularisation potential as the deviation to the reference RDFs is reducing. Finally, it is worth mentioning a possible future collaboration with Prof. Lars Nordenskiöld from Nanyang Technological University in Singapore who is working both experimentally and theoretically on DNA compaction and chromatine condensation.

6 Projected publications/articles resulting or to result from the grant

The results presented here concerning the IMC approach will be part of an invited article in a themed issue on multi-scale modelling in Physical Chemistry Chemical Physics to be published in 2011. A first manuscript on the application of the MARTINI procedure to DNA is planned in the beginning of 2011.

References

[1] Voth, G. A. *Coarse-graining of condensed phase and biomolecular systems*. Taylor & Francis Group, New York, 2009.

[2] Tepper, H. L., Voth, G. A. A coarse-grained model for double-helix molecules in solution: spontaneous helix formation and equilibrium properties. *J. Chem. Phys.*, **2005**, 122, 124906.

[3] Knotts, T. A., Rathore, N., Schwartz, D. C., de Pablo, J. J. A coarse grain model for DNA. *J. Chem. Phys.*, **2007**, 126, 084901.

[4] Becker, N. B., Everaers, R. From rigid base pairs to semi-flexible polymers: coarse-graining DNA. *Phys. Rev. E.: Stat. Nonlin. Soft. Matter Phys.*, **2007**, 76, 021923.

[5] Poulain, P., Saladin, A., Hartmann, B., Prévost, C. Insights on protein-DNA recognition by coarse grain modeling. *J. Comput. Chem.*, **2008**, 29, 2582-2592.

[6] Khalid, S., Bond, P. J., Holyoake, J., Hawtin, R. W., Sansom, M. S. P. DNA and lipid bilayers: self-assembly and insertion. *J. R. Soc. Interface*, **2008**, 5, S241-S250.

[7] Dans, P. D., Zeida, A., Machado, M. R., Pantano, S. A coarse grained model for atomic-detailed simulations with explicit electrostatics. *J. Chem. Theory Comput.*, **2010**, 6, 1711-1725.

[8] Marrink, S. J., de Vries, A. H., Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B*, **2004**, 108, 750-760.

[9] Marrink, S. J., Risselada, H. J., Yefimov, S., Tielkeman, D. P., de Vries, A. H. The MARTINII force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, **2007**, 111, 7812-7824.

[10] Monticelli, L., Kandasamy, S. K., Periole, X., Larson, R. G., Tieleman, D. P., Marrink, S. J. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.*, **2008**, 4, 819-834.

[11] López, C. A., Rzepiela, A. J., de Vries, A. H., Dijkhuizen, L., Hunenberger, P. H., Marrink, S. J. Martini coarse-grained force field: extension to carbohydrates. *J Chem. Theory Comput.*, **2009**, 5, 3195-3210.

[12] Henderson, R. L. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A*, **1974**, 49, 197-198.

[13] Lyubartsev, A. P., Laaksonen, A. Calculation of effective potentials from radial distribution functions: a reverse Monte Carlo approach. *Phys. Rev. E*, **1995**, 52, 3730-3737.
[14] Lyubartsev, A. P., Laaksonen, A. Osmotic and activity coefficients from effective potentials for hydrated ions. *Phys. Rev. E*, **1997**, 55, 5689-5696.
[15] Lyubartsev, A. P., Laaksonen, A. Effective potentials for ion-DNA interactions. *J. Chem. Phys.*, **1999**, 111, 11207-11215.
[16] Lyubartsev, A. P., Laaksonen, A. Determination of effective pair potentials from ab initio simulations: application to liquid water. *Chem. Phys. Lett.*, **2000**, 325, 15-25.
[17] Lyubartsev, A. P., Laasonen, K., Laaksonen, A. Hydration of Li^+ ion. An ab initio molecular dynamics simulation. *J. Chem. Phys.*, **2001**, 114, 3120-3126.
[18] Lyubartsev, A. P., Marčelja, S. Evaluation of effective ion-ion potentials in aqueous electrolytes. *Phys. Rev. E*, **2002**, 65, 041202.
[19] Lyubartsev, A. P. Multiscale modeling of lipids and lipid bilayers. *Eur. Biophys. J.*, **2005**, 35, 53-61.
[20] Lyubartsev, A. P., Mirzoev, A., Chen, L., Laaksonen, A. Systematic coarse-graining of molecular models by the Newton inversion method. *Faraday Discuss.*, **2010**, 144, 43-56.
[21] Lavery, R., Zakrzewska, K., Beveridge, D., Bishop, T. C., Case, D. A., Cheatham, T. III, Dixit, S., Jayaram, B., Lankas, C., Laughton, J. H., Maddocks, A., Michon, R., Osman, F., Orozco, M., Perez, A., Singh, T., Spackova, N., Sponer, J. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucl. Acids Res.*, **2010**, 38, 299-313.
[22] Naômé, A., Schyman, P., Laaksonen, A., Vercauteren, D. P. Molecular dynamics simulation of 8-oxoguanine containing DNA fragments reveals altered hydration and ion binding patterns. *J. Phys. Chem. B*, **2010**, 114, 4789-4801
[23] Lyubartsev, A. P., Laaksonen, A. MDynaMix – a scalable portable parallel MD simulation package for arbitrary molecular mixtures. *Comput. Phys. Comm.*, **2000**, 128, 565-589.
[24] Soper, A.K. Tests of the empirical potential structure refinement method and a new method of application to neutron diffraction data on water. *Mol. Phys.*, **2001**, 99, 1503-1516.
[25] Cullis, P. M., Wolfenden, R. Affinities of nucleic acid bases for solvent water. *Biochemistry*, **1981**, 20, 3024-3028.
[26] Shih, P., Pedersen, L. G., Gibbs, P. R., Wolfenden, R. Hydrophobicities of the nucleic acid bases: distribution coefficients from water to cyclohexane. *J. Mol. Biol.*, **1998**, 280, 421-430.
[27] Oostenbrink, C., Villa, A., Mark, A. E., van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force parameter sets 53A5 and 53A6. *J. Comput. Chem.*, **2004**, 25, 1656-1676.
[28] Berendsen, H. C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*. Pullman, B., Ed., Dordrecht, Reidel, **1981**, pp 331-342.
[29] Tironi, I. G., van Gunsteren, W. F. A molecular dynamics simulation study of chloroform. *Mol. Phys.*, **1994**, 83, 381-403.
[30] Hess, B., Kutzner, C., vand der Spoel, D., Lindahl, E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **2008**, 4, 435-447.
[31] Young, P. E., Hillier, I. H. Hydration free energies of nucleic acid bases using an ab initio continuum model. *Chem. Phys. Lett.*, **1993**, 215, 405-408.
[32] Young, P. E., Hillier, I. H., Gould, I. R. Accurate prediction of the solvation of nucleotide base pairs using an ab initio continuum model. *J. Chem. Soc. Perkin Trans. 2*, **1994**, 1717-1718.
[33] Orozco, M., Colominas, C., Luque, F. J., Theoretical determination of the solvation free energy in water and chloroform of the nucleic acid bases. *Chem. Phys.*, **1996**, 209, 19-29.
[34] Giesen, D. J., Chambers, C. C., Cramer, C. J., Truhlar, D. G. What controls partitioning of the nucleic acid bases between chloroform and water? *J. Phys. Chem. B*, **1997**, 101, 5084-5088.
[35] Eksterowicz, J. E., Miller, J. L., Kollman, P. A. Calculation of chloroform/water partition coefficients for N-methylated nucleic acid bases. *J. Phys. Chem. B*, **1997**, 101, 10971-10975.
[36] Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E. III, Laughton, C. A., Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **2007**, 92, 3817-3829.
[37] Dang, L. X. Mechanism and thermodynamics of ion selectivity in aqueous-solutions of 18-crown-6 ethers: a molecular dynamics study. *J. Am. Chem. Soc.*, **1995**, 117, 6954-6960.