Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Applications Note

Structural bioinformatics

Integrative structural and dynamical biology with PLUMED-ISDB

Massimiliano Bonomi[1,*] and Carlo Camilloni[2,3,*]

[1]Department of Chemistry, University of Cambridge, Address Lensfield Road, CB2 1EW Cambridge, *UK*., [2]Department of Chemistry and Institute for Advanced Study, Technische Universität München, Address Lichtenbergstr. 4, 85747 Garching, Germany. [3]Dipartimento di Bioscienze, Università degli Studi di Milano, 20133 Milan, Italy

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: Accurate structural models of biological systems can be obtained by properly combining experimental data with *a priori* physico-chemical knowledge. Here we present PLUMED-ISDB, an open-source, freely-available module of the popular PLUMED library, which enables the simultaneous determination of structure and dynamics of conformationally heterogeneous systems by integrating experimental data with *a priori* information. This integration is achieved using metainference, a general Bayesian framework that accounts for both noise in the data and their ensemble-averaged nature. PLUMED-ISDB implements different types of experimental data, such as several NMR observables, FRET, SAXS and cryo-electron microscopy data, and enables modelling structure and dynamics of individual proteins, protein complexes, membrane proteins, RNA, and DNA, using a variety of enhanced sampling methods and resolutions of the system.

Availability and Implementation: PLUMED-ISDB is freely available on the web as part of the PLUMED library at www.plumed.org

Contact: Massimiliano Bonomi: mb2006@cam.ac.uk, Carlo Camilloni: carlo.camilloni@unimi.it

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

To uncover the mechanisms used by macromolecular machines to perform their biological functions, it is critical to characterize both their structure and dynamics. Single experimental techniques struggle to achieve alone this goal, mostly because data typically provide sparse information affected by random and systematic errors. Furthermore, in complex biological systems that adopt multiple conformations, experimental observations depend on an ensemble of states, making the contribution to the data of each member of the ensemble hard to disentangle. On the other hand, theoretical and computational approaches based on physics, chemistry or statistics are not accurate enough to predict *ab initio* structure and dynamics of a complex system. Integrative methods that properly combine experimental data with *a priori* knowledge have the potential to accurately characterize the behavior of complex systems. A number of different integrative methods have been proposed over the years (Bonomi, et al., 2017). These methods are built around a series of different components. First, a predictor of the experimental observable is required to quantify the consistency of the modelled ensemble with the input data. Second, a framework to treat simultaneously conformational heterogeneity and errors in the data is needed as all biological systems and experiments almost present both. Third, efficient generation of structural models requires advanced sampling techniques, beyond the capabilities of standard Molecular Dynamics (MD) or Monte Carlo. Finally, biological systems span different length scales and thus should be represented at a resolution appropriate to the available information. Here, we present PLUMED-ISDB, a module of the popular PLUMED library (Tribello, et al., 2014) for Integrative Structural and Dynamical Biology. PLUMED-ISDB provides a computational toolbox that enables using several types of experimental data to simultaneously model structure and dynamics of biological systems. PLUMED-ISDB is fully integrated with PLUMED, giving access to many enhanced sampling techniques and representations of the system, from atomistic to coarse-grained level. A detailed description of the methods available can be found in the on-line manual (https://plumed.github.io/doc-isdb/user-doc/html/index.html) and in the SI. Starting from PLUMED version 2.4, this documentation will be part of the official distribution and accessible at www.plumed.org.

2 Algorithms

2.1 Metainference

Metainference (Bonomi, et al., 2016) is a Bayesian framework (Rieping, et al., 2005) to model heterogeneous systems by integrating prior information with noisy, ensemble-averaged data. Metainference models a system and quantifies the level of noise in the data by considering a set of replicas of the system. PLUMED-ISDB implements multiple models of data error, including Gaussian and outliers noise (Bonomi, et al., 2016), and methods to automatically determine all the metainference parameters, such as the statistical uncertainty in modelling a finite ensemble, data scaling factor and offset (Löhr, et al., 2017). The metainference approach can be used in combination with all the experimental data types described in the next paragraph. An example application is presented in Fig. 1.

.

2.2 Experimental data

Observable	Description
Nuclear Overhauser Effect	Isolated spin-pair approximation
Proteins chemical shifts	Camshift (Kohlhoff, et al., 2009)
^{3}J-couplings	Karplus' equations (Vogeli, et al., 2007);
Residual dipolar couplings and dipolar couplings	Theta method (Camilloni and Vendruscolo, 2015)
Pseudocontact shifts	Theta method (Camilloni and Vendruscolo, 2015)
Paramagnetic resonance enhancement	Solomon-Blomberg approximation (Iwahara, et al., 2004)
Förster resonance energy transfer	Förster equation
Small-angle X-ray scattering	Debye formula with solvent effects (Fraser, et al., 1978)
Cryo-electron microscopy	Bayesian treatment of cryo-EM data (Hanot, et al., 2017)

3 Software

PLUMED-ISDB is distributed as module of the PLUMED library and provides the functionalities described in the previous sections. A dedicated input file should be prepared to specify, in the format of PLUMED directives, the experimental data used in the modelling and the metainference setup. Additional functional forms of experimental data not natively supported can be defined directly in the input file. PLUMED-ISDB can be used in combination with several MD codes, state-of-the-art atomistic and coarse-grained force fields, and enhanced sampling techniques, via the PLUMED library. The combination of metainference and metadynamics (Bonomi, et al., 2016) is natively supported. PLUMED-ISDB can be used for single structure refinement, ensemble determination, or to calculate new experimental observables on pre-existing models using the PLUMED *driver* utility.

4 Software library and implementation

PLUMED-ISDB is written in C++ and parallelized using MPI and openMP. The module is freely available and distributed along with PLUMED at www.plumed.org. The development version can be accessed on GitHub (github.com/plumed/plumed2/tree/isdb). A series of examples and tests, currently available in the isdb branch, will be included in the official PLUMED distribution starting from version 2.4. For discussions, the user can refer to plumed-users@googlegroups.com.

Funding

CC was supported by the Technische Universität München – Institute for Advanced Study, funded by the German Excellence Initiative and the EU Seventh Framework Programme under grant agreement n° 291763.

Conflict of Interest: none declared.

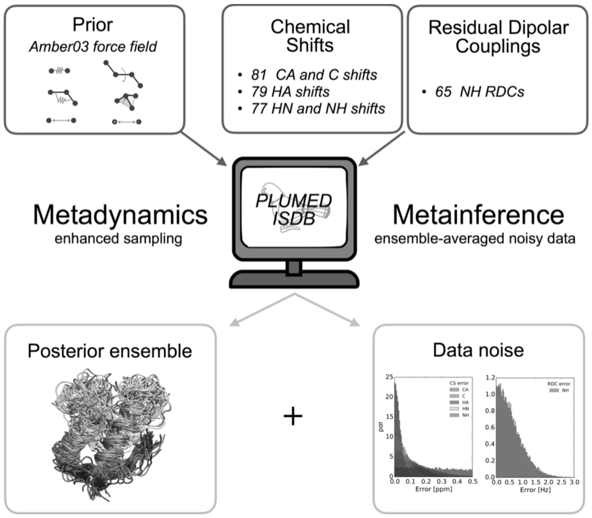

Figure 1. Application of the Metadynamic Metainference approach (Bonomi, et al., 2016) to the Colicin Immunity protein IM7. Chemical Shifts and Residual Dipolar Couplings NMR data are integrated with the Amber03 force field to determine the posterior ensemble of models and the level of noise in the input data.

References

Bonomi, M., *et al.* Metainference: a Bayesian inference method for heterogeneous systems. *Sci. Adv.* 2016;2:e1501177.

Bonomi, M., Camilloni, C. and Vendruscolo, M. Metadynamic metainference: Enhanced sampling of the metainference ensemble using metadynamics. *Sci. Rep.* 2016;6:31232.

Bonomi, M., *et al.* Principles of protein structural ensemble determination. *Curr Opin Struct Biol* 2017;42:106-116.

Camilloni, C. and Vendruscolo, M. A tensor-free method for the structural and dynamical refinement of proteins using residual dipolar couplings. *J Phys Chem B* 2015;119(3):653-661.

Camilloni, C. and Vendruscolo, M. Using Pseudocontact Shifts and Residual Dipolar Couplings as Exact NMR Restraints for the Determination of Protein Structural Ensembles. *Biochemistry-Us* 2015;54(51):7470-7476.

Fraser, R.D.B., Macrae, T.P. and Suzuki, E. Improved Method for Calculating Contribution of Solvent to X-Ray-Diffraction Pattern of Biological Molecules. *J Appl Crystallogr* 1978;11(Dec):693-694.

Hanot, S., *et al.* Multi-scale Bayesian modeling of cryo-electron microscopy density maps. *bioRxiv* 2017;doi: 10.1101/113951.

Iwahara, J., Schwieters, C.D. and Clore, G.M. Ensemble approach for NMR structure refinement against H-1 paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J Am Chem Soc* 2004;126(18):5879-5896.

Kohlhoff, K.J., *et al.* Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J Am Chem Soc* 2009;131(39):13894-13895.

Löhr, T., Jussupow, A. and Camilloni, C. Metadynamic metainference: Convergence towards force field independent structural ensembles of a disordered peptide. *The Journal of Chemical Physics* 2017;146(16):165102.

Rieping, W., Habeck, M. and Nilges, M. Inferential structure determination. *Science* 2005;309(5732):303-306.

Tribello, G.A., *et al.* PLUMED 2: New feathers for an old bird. *Comp. Phys. Comm.* 2014;185(2):604-613.

Vogeli, B., *et al.* Limits on variations in protein backbone dynamics from precise measurements of scalar couplings. *J Am Chem Soc* 2007;129(30):9377-9385.