short communications

NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps

Karsten Suhre,[a]* Jorge Navaza[b] and Yves-Henri Sanejouand[c]

[a]Genomics and Structural Information Laboratory, Institut de Biologie Structurale et Microbiologie, UPR 2589 CNRS, Marseille, France, [b]LMES, Institut de Biologie Structurale Jean-Pierre Ebel, UMR 5075 CEA–CNRS–Universite Joseph Fourier, Grenoble, France, and [c]Laboratoire de Physique, UMR 5672 of CNRS, Ecole Normale Supérieure, Lyon, France

Correspondence e-mail:
karsten.suhre@igs.cnrs-mrs.fr

This paper describes a freely available software suite that allows the modelling of large conformational changes of high-resolution three-dimensional protein structures under the constraint of a low-resolution electron-density map. Typical applications are the interpretation of electron-microscopy data using atomic scale X-ray structural models. The software package provided should enable the interested user to perform flexible fitting on new cases without encountering major technical difficulties. The NORMA software suite including three fully executable reference cases and extensive user instructions are available at http://www.elnemo.org/NORMA/.

1. Introduction

High-quality three-dimensional reconstructions of large protein assemblies are increasingly becoming available owing to recent advances in cryo-electron microscopy (EM) and related technologies. In many cases, atomic scale models of the involved proteins are determined in parallel by X-ray crystallography. It is then possible to attempt to fit these high-resolution (X-ray) models into the low-resolution EM density map. However, it turns out that these proteins sometimes do not fit well into the EM reconstruction, indicating that they adopt a significantly different conformation in their 'functional' (often multimeric) environment to that under crystallization conditions. A determination of the related conformational changes can then not only help in the interpretation of the EM observations, but in addition yield valuable information about the functional movements of the proteins involved.

It has been shown in the past that large conformational changes often correspond to highly collective movements that can be described well by a small number of low-frequency normal modes of that protein (Harrison, 1984; Krebs et al., 2002; Marques & Sanejouand, 1995; McCammon et al., 1976; Perahia & Mouawad, 1995; Tama & Sanejouand, 2001; Yang & Bahar, 2005). It was thus not overly surprising to find that normal-mode perturbed models could be used to phase X-ray diffraction data by molecular replacement (Suhre & Sanejouand, 2004a) and also in the subsequent model-refinement step (Delarue & Dumas, 2004). More recently, the application of normal-mode flexible fitting into EM density maps has been formally introduced by Tama et al. (2004a). Three prominent examples where this approach yielded exciting results in a biologically relevant context are the membrane protein CaATPase (Hinsen et al., 2005), the protein-conducting channel bound to a translating ribosome (Mitra et al., 2005) and a chaperonin GroEL–protein substrate complex (Falke et al., 2005).

Clearly, flexible and freely available software tools are key to opening this exciting field of research to a wider community. This has been documented by the success of the recent EMBO Practical Course on Combination of Electron Microscopy and X-ray Crystallography in Structure Determination, which was held in October 2005 in Gif-sur-Yvette, France. In this paper we present a software suite named NORMA, which originates from initial developments for this EMBO workshop, and is based on a combination of earlier work from the authors on the EM-density fitting program URO (Navaza et al.,

2002) and on a normal-mode (NMA) code that is implemented in the web server *elNémo* (Suhre & Sanejouand, 2004*b*; http://www.elnemo.org/).

2. The *NORMA* software suite

URO (Navaza *et al.*, 2002) is a fast method for fitting protein structural models into EM reconstructions. The methodology is inspired by the molecular-replacement technique (Navaza, 2001), adapted to take into account phase information and the symmetry imposed during the EM reconstruction. Calculations are performed in reciprocal space, which enables the selection of large volumes of the EM maps, thus avoiding the bias introduced when defining the boundaries of the target density. *elNémo* (Suhre & Sanejouand, 2004*a,b*) is a rapid NMA code that implements two major approximations which allow the computation of the lowest frequency normal modes for large protein complexes at an all-atom level of description. These are the elastic network approximation (Tirion, 1996) and the building-block approach (RTB method; Durand *et al.*, 1994; Li & Cui, 2002; Tama *et al.*, 2000).

In practice, to perform the flexible fitting a set of amplitudes needs to be determined for a given number of low-frequency normal modes that minimize the *URO* misfit parameter (Q), where the corresponding NMA-perturbed model is used as a search model. Q is the normalized quadratic misfit between the electron density of the model and the EM map. For this task, a multiple-dimension simplex-minimization algorithm with optional simulated annealing has been chosen (Press *et al.*, 1992). A set of shell scripts has been developed to couple *URO* and *elNémo via* the minimizer. The resulting software suite, named *NORMA*, has been applied to three reference cases: (i) fitting of the open conformation of the GroEL chaperonin into an EM map of the closed conformation, (ii) optimization of the fit of the major IBDV capsid protein and (iii) modelling of the conformational change of the structure of the CaATPase molecule between its isolated form and its membrane-bound form. The latter case has been used as a benchmark, since the authors of the original work (Hinsen *et al.*, 2005) used *URO* to score the CaATPase fitting, while their fitting algorithm is distinctively different from that which we use here (direct-space fitting and a different normal-mode analysis approach). We show that *NORMA* is able to closely reproduce these results in terms of misfit parameter Q and amplitudes of the major excited normal modes.

The entire software package, including *URO* and *elNémo*, has been made freely available over the web at http://www.elnemo.org/NORMA/. This site provides detailed installation instructions including a user guide, fully configured data sets for the three reference cases and example results. Different fitting protocols are proposed and discussed. The provided software package should enable the interested user to perform flexible fitting on new cases without encountering major technical difficulties. Technical details are presented (and will evolve in the future) on the *NORMA* web site and will not be discussed here.

3. Flexible fitting of GroEL with *NORMA*: an example

Large domain movements are critical to GroEL-mediated protein folding. This chaperonin has been extensively studied in the past. High-resolution models are available for the open (Ranson *et al.*, 2001; PDB code 1aon) and for the closed (Chaudhry *et al.*, 2004; PDB code 1sx3) conformation of a single GroEL molecule. De Carlo *et al.* (2002) have determined a cryo-electron microscopy reconstruction of the GroEL chaperonin complex in its 'closed form'. We can thus pose the hypothetical challenge of determining the closed form of a GroEL molecule at atomic resolution based solely on the high-resolution X-ray structure of its open conformation and the low-resolution EM reconstruction of its closed form. Such a scenario corresponds to a typical application of normal-mode fitting with a tool such as *NORMA*. The structure of the closed form of GroEL will only serve as a reference. Details of this test case, provided in a completely reproducible form that runs on a standard Linux PC, are given on the *NORMA* web site. The results can be summarized as follows.

55% of the conformational change between 1aon and 1sx3 can be explained by a movement that follows the lowest frequency normal mode of 1aon (as computed using the *elNémo* web server; link provided on the *NORMA* web site). Consequently, normal-mode fitting with *NORMA* using only the single lowest frequency mode of 1aon already yields relatively good results. The root-mean-square distance (r.m.s.d.) between the one-mode fitted model and the reference structure 1sx3 (closed form) is only 7.9 Å, compared with an r.m.s.d. of 12 Å between the original structures. The correlation coefficient increases from 0.615 for the unperturbed (open) conformation to 0.760 for the one-mode fitted structure. This value should be compared with a maximal obtainable correlation coefficient of 0.853
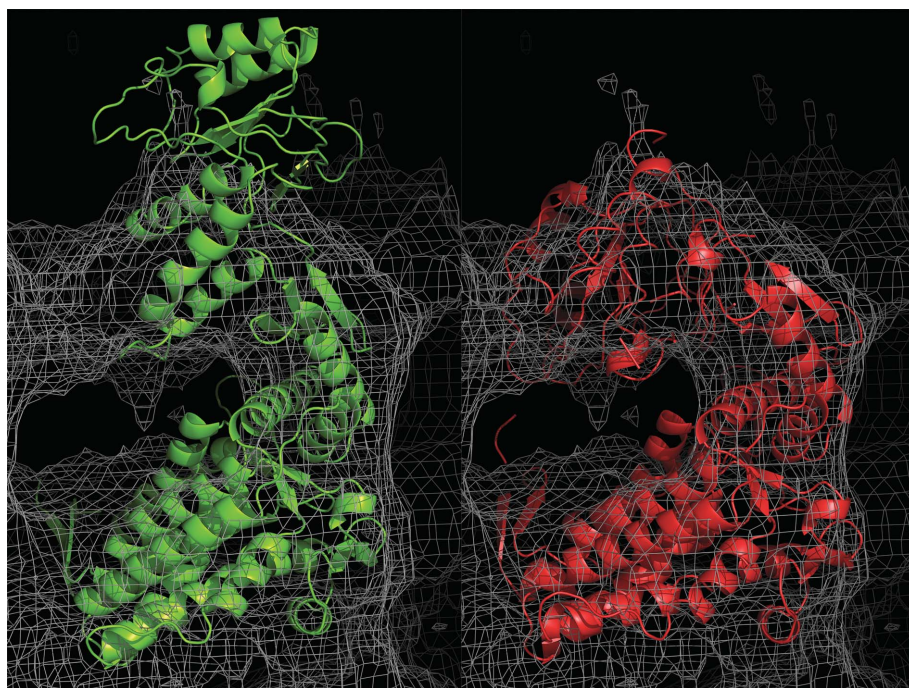

Figure 1
Chaperonine GroEL fitted to the EM reconstruction (left, unperturbed 1aon structure fitted using *URO*; right, normal-mode perturbed 1aon structure using *NORMA* in a four-step approach and intermediate model regularization). Note that only one of the 14 fitted copies of the GroEL molecule is shown here. This figure was prepared with *PyMOL* (DeLano, 2002).

that is reached when using the (closed) reference structure. *NORMA* fitting with five modes further decreases the r.m.s.d. with respect to 1sx3 to 5.8 Å; the correlation coefficient rises to a value of 0.788. However, a further increase of the number of modes used in the flexible fitting eventually leads to a situation of over-fitting: although the correlation coefficient for a ten-mode fitting increases to a value of 0.833, the r.m.s.d. with respect to the reference structure also increases to attain a value of 9.3 Å.

However, visual inspection of the fitting process (see animations on the *NORMA* web site) suggests that such a large conformational change is best broken up into a number of smaller steps. We therefore computed in the first step the optimal fit using the lowest frequency mode of GroEL with *NORMA*, but we applied only 30% of the corresponding amplitude to generate a first intermediate model. This step was completed by a model regularization using *REFMAC* (Murshudov *et al.*, 1997) to 'repair' inevitable bond distortions that are induced when applying large normal-mode perturbations to a protein. The resulting model was then used to initiate the second very similar step, where we now applied 50% of the computed optimal perturbation. A third step followed where 100% of the perturbation was applied. In the fourth and final fitting step, a flexible fitting with 12 low-frequency normal modes was then used to allow more localized protein deformations in order to best fit the EM density. The result of this multi-step approach is shown in Fig. 1. An animation is available on the *NORMA* web site.

4. Concluding remarks

The objective of this paper is to make the technique of flexible fitting of protein models into density maps from EM reconstructions accessible to a wider community of crystallographers and structural biologists. Although we believe that the choices made in *NORMA* are optimal for our purpose, alternative and/or complementary programs that have been developed by other authors can be implemented with little effort within the *NORMA* scripting framework. For example, Wriggers and coworkers developed the real-space fitting program *SITUS* (Wriggers & Birmanns, 2001; Wriggers *et al.*, 1999), which represents an alternative to the reciprocal-space fitting method *URO*. Tama *et al.* (2004b) have used a minimization algorithm that optimizes the normal-mode amplitudes one by one in an iterative manner. This approach may be more economical in terms of computing time, but has a higher chance of becoming stuck in a local minima. Alternative minimization tools can also be found in Press *et al.* (1992).

When large conformational changes are involved, it may also be essential to apply intermediate structure-regularization steps to keep bond lengths and torsion angles within reasonable bounds. Such steps can also be easily implemented in the *NORMA* scripts, as exemplified by the GroEL reference case (see *NORMA* web site). A generalization to multi-protein fitting problems is straightforward and only requires minor adaptation of the *URO* input parameters and duplication of calls to the NMA package. *NORMA* has been developed

and extensively tested on different Linux implementations. Portability to other Unix platforms should be simple and will be supported by the authors.

References

Chaudhry, C., Horwich, A. L., Brunger, A. T. & Adams, P. D. (2004). *J. Mol. Biol.* **342**, 229–245.
De Carlo, S., El-Bez, C., Alvarez-Rua, C., Borge, J. & Dubochet, J. (2002). *J. Struct. Biol.* **138**, 216–226.
DeLano, W. L. (2002). *The PyMOL Molecular Visualization System.* DeLano Scientific, San Carlos, CA, USA.
Delarue, M. & Dumas, P. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 6957–6962.
Durand, P., Trinquier, G. & Sanejouand, Y. H. (1994). *Biopolymers*, **34**, 759–771.
Falke, S., Tama, F., Brooks, C. L. III, Gogol, E. P. & Fisher, M. T. (2005). *J. Mol. Biol.* **348**, 219–230.
Harrison, W. (1984). *Biopolymers*, **23**, 2943–2949.
Hinsen, K., Reuter, N., Navaza, J., Stokes, D. L. & Lacapere, J. J. (2005). *Biophys. J.* **88**, 818–827.
Krebs, W. G., Alexandrov, V., Wilson, C. A., Echols, N., Yu, H. & Gerstein, M. (2002). *Proteins*, **48**, 682–695.
Li, G. & Cui, Q. (2002). *Biophys. J.* **83**, 2457–2474.
McCammon, J. A., Gelin, B. R., Karplus, M. & Wolynes, P. G. (1976). *Nature (London)*, **262**, 325–326.
Marques, O. & Sanejouand, Y. H. (1995). *Proteins*, **23**, 557–560.
Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks, C. L. III, Ban, N. & Frank, J. (2005). *Nature (London)*, **438**, 318–324.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Navaza, J., Lepault, J., Rey, F. A., Alvarez-Rua, C. & Borge, J. (2002). *Acta Cryst.* D**58**, 1820–1825.
Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1992). Editors. *Numerical Recipes In Fortran77: The Art Of Scientific Computing.* Cambridge University Press.
Perahia, D. & Mouawad, L. (1995). *Comput. Chem.* **19**, 241–246.
Ranson, N. A., Farr, G. W., Roseman, A. M., Gowen, B., Fenton, W. A., Horwich, A. L. & Saibil, H. R. (2001). *Cell*, **107**, 869–879.
Suhre, K. & Sanejouand, Y. H. (2004a). *Acta Cryst.* D**60**, 796–799.
Suhre, K. & Sanejouand, Y. H. (2004b). *Nucleic Acids Res.* **32**, W610–W614.
Tama, F., Gadea, F. X., Marques, O. & Sanejouand, Y. H. (2000). *Proteins*, **41**, 1–7.
Tama, F., Miyashita, O. & Brooks, C. L. III (2004a). *J. Mol. Biol.* **337**, 985–999.
Tama, F., Miyashita, O. & Brooks, C. L. III (2004b). *J. Struct. Biol.* **147**, 315–326.
Tama, F. & Sanejouand, Y. H. (2001). *Protein Eng.* **14**, 1–6.
Tirion, M. M. (1996). *Phys. Rev. Lett.* **77**, 1905–1908.
Wriggers, W. & Birmanns, S. (2001). *J. Struct. Biol.* **133**, 193–202.
Wriggers, W., Milligan, R. A. & McCammon, J. A. (1999). *J. Struct. Biol.* **125**, 185–195.
Yang, L. W. & Bahar, I. (2005). *Structure*, **13**, 893–904.