PiSQRD: a web server for decomposing proteins into quasi-rigid dynamical domains

T. Aleksiev¹, R. Potestio¹,*-‡, F. Portiggia¹, S. Cozzini¹,² and C. Micheletti¹,²,³

¹Scuola Internazionale Superiore di Studi Avanzati and eLab, ²CNR-INFM Democritos Simulation Centre, via Beirut 2-4, 34151 Trieste and ³Italian Institute of Technology, Genoa, Italy

Received on June 17, 2009; revised on August 10, 2009; accepted on August 16, 2009

Advance Access publication August 20, 2009

ABSTRACT

Summary: The PiSQRD web resource can be used to subdivide protein structures in quasi-rigid dynamical domains. The latter are groups of amino acids behaving as approximately rigid units in the course of protein equilibrium fluctuations. The PiSQRD server takes as input a biomolecular structure and the desired fraction of protein structural fluctuations that must be accounted for by the relative rigid-body motion of the dynamical domains. Next, the lowest energy modes of fluctuation of the protein (optionally provided by the user) are calculated and used to identify the rigid subunits. The resulting optimal subdivision is returned through a web page containing both interactive graphics and detailed data output.

Availability: The PiSQRD web server, which requires Java, is available free of charge for academic users at http://pisqrd.escience-lab.org.

Contact: potestio@sissa.it

1 INTRODUCTION

Concerted internal movements are ubiquitous in proteins and often play a key role in assisting or accompanying the biological functionality of these molecules. The lowest energy modes of protein structural fluctuations, those that are most easily excited by thermal energy, typically have a collective character, in that groups of several amino acids move coherently. This fact suggests the possibility to describe protein internal fluctuations in terms of the relative motion of quasi-rigid subunits.

A variational scheme to decompose proteins in quasi-rigid domains on the basis of the low-energy fluctuation modes was recently introduced by some of us in (Potestio et al., 2009). The method aptly complements in several important respects other existing tools for identifying protein quasi-rigid domains. Available web servers include Dyndom and TLSMD (Hayward et al., 1997; Painter and Merritt, 2006). Dyndom subdivides single-chain proteins based on large-scale deformations calculated from alternative protein conformers obtained experimentally or modelled computationally. The latter scheme is used, for example, by ProMode which offers a database of precalculated dynamical domains for selected PDB chains (Wako et al., 2004). Conversely, TLSMD applies the translation–libration–screw-like motion analysis of Schomaker and Trueblood (1968) to crystallographic B-factor measurements so as to identify groups of amino acids whose fluctuation dynamics is compatible with that of a rigid body. To limit the computational demand, each TLSMD domain is constrained to span an uninterrupted stretch of the primary sequence. In addition, we mention the availability of standalone programs such as Domain Finder (Hinsen, 2000) where individual low-energy modes can be used to identify quasi-rigid blocks of amino acids.

In this work, we describe the implementation of the algorithm of Potestio et al. (2009) as a web server, termed PiSQRD after protein structure quasi-rigid domain decomposition, where users can submit queries for subdividing proteins into quasi-rigid units. In comparison with the aforementioned methods the server: (i) can decompose protein complexes consisting of several peptide chains; (ii) makes use of an entire set of low-energy modes calculated ‘on-the-fly’ via an elastic-network model or uploaded by the user; and (iii) allows each quasi-rigid domain to gather amino acids covering different stretches of the primary sequence (indeed, in multimeric entries, domains can even span different protein chains). The viability/quality of the returned subdivision is conveyed by a subdivision score. The latter measures the fraction of protein internal dynamics captured after the suppression of internal structural fluctuations in the quasi-rigid units.

The main assumption on which the approach relies is that the considered low-energy modes correspond to protein fluctuations around a well-defined reference structure (as opposed to large-scale ‘jumps’ across structurally heterogeneous substrates). Applications for the method range from the analysis and rationalization of protein internal dynamics data obtained from experiments or molecular dynamics simulations to advanced stochastic techniques where the sampling of protein conformational substates is enhanced/accelerated by the relative rigid-like displacements of the rigid domains.

The PiSQRD web server is freely accessible and accepts submissions of subdivision queries that can be evaluated within a preassigned CPU time of 20 min. The computational time required by the algorithm grows almost quadratically as a function of the sequence length. Decomposing a protein of about 200 amino acids (as 4akeA in the example of (Fig. 1), 214 amino acids) requires ~50 s. The CPU time limit is thus sufficient to decompose proteins of up to 800 amino acids. Academic users planning to run more computationally intensive subdivisions can register (free of charge) to have access to a HPC and GRID infrastructure.

*To whom correspondence should be addressed.
The PiSQRD web tool is presented to the user as a single-page input form. The algorithm introduced in Potestio et al. (2009) and implemented in PiSQRD relies on the observation that the distinctive property of a rigid body is that pairwise distances between any two of its constitutive points remain unaltered during its motion. The structural deformations described by the low-energy modes can be aptly used to quantify the extent to which the pairwise distances of a set of amino acids remain constant during protein equilibrium fluctuations. This measure provides a quantitative criterion to assess whether a certain grouping of amino acids is a viable decomposition of the structure into quasi-rigid units.

The method performs a stochastic search in the space of possible amino acid groupings to identify the best partitioning of amino acids into a number of quasi-rigid domains, \( Q \), ranging from 2 to 20. For each value of \( Q \) the best assignment of the amino acids to the \( Q \) domains is found by maximizing the quantity \( f \), which conveys the fraction of protein internal essential dynamics captured by the rigid units decomposition.

\[
    f = \sum_{l=1}^{n} \left( \frac{\lambda_l}{\lambda_1} \right)^2
\]

where \( \bar{\lambda} \) is the rigid-body approximation to the \( l \)-th mode obtained by suppressing the internal deformations of each of the \( Q \) units; \( \lambda_l \) is the mean square protein fluctuation captured by the \( l \)-th mode and \( n \) is the number of used low-energy modes. Typically, the first \( n=10 \) modes, also termed essential modes, are sufficient to account for most of the protein fluctuations. Since \( [\bar{\lambda}]^2 \leq 1 \), then \( f \) takes on values in the [0, 1] range. A perfect (very poor) rigid body decomposition is associated with \( f = 1 \) (\( f = 0 \)). The optimized \( f \) is, expectedly, a non-decreasing function of \( Q \).

\section*{3 IMPLEMENTATION}

The PiSQRD web server performs an automatic subdivision of proteins into quasi-rigid domains. The inputs are the protein structure input file which should at least contain the protein’s Cα trace. Unless a specific subchain of interest is specified by the user in an optional input field, the subdivision is performed on the complete structure file (which may comprise several chains). Finally, the user indicates the desired threshold fraction of internal essential dynamics, \( f \). The first \( n \) energy essential modes are next calculated and are used to decompose the protein from \( Q=2 \) to 20 quasi-rigid domains. The returned optimal decomposition is the one having the smallest \( Q \) that is sufficient to capture the pre-assigned threshold of internal essential mobility (three domains in the example shown, using the default values: \( f = 80\% \), \( n=10 \)).

- a histogram with the (size of amino acids) of each domain in the best decomposition;
- a one-dimensional sequence representation of the best domain decomposition; and
- links to detailed graphical/data output to all subdivisions from 2 to 20 domains.

\section*{4 CONCLUDING REMARKS}

The PiSQRD web server performs an automatic subdivision of proteins into quasi-rigid domains. The inputs are the protein structure file (4akeA in the example), and the desired threshold fraction of internal essential dynamics, \( f \). The first \( n \) lowest energy essential modes are next calculated and used to decompose the protein from \( Q=2 \) to 20 quasi-rigid domains. The returned optimal decomposition is the one having the smallest \( Q \) that is sufficient to capture the pre-assigned threshold of internal essential mobility (three domains in the example shown, using the default values: \( f = 80\% \), \( n=10 \)).

\begin{figure}[h]
    \centering
    \includegraphics[width=\textwidth]{figure1.png}
    \caption{Graphical summary of the PiSQRD flowchart. Users provide the input structure (4akeA in the example), and the desired threshold fraction of internal essential dynamics, \( f \). The \( n \) lowest energy essential modes are next calculated and used to decompose the protein from \( Q=2 \) to 20 quasi-rigid domains. The returned optimal decomposition is the one having the smallest \( Q \) that is sufficient to capture the pre-assigned threshold of internal essential mobility (three domains in the example shown, using the default values: \( f = 80\% \), \( n=10 \)).}
\end{figure}

\section*{REFERENCES}
