Integration of FIRST, FRODA and NMM in a coarse grained method to study Protein Disulphide Isomerase conformational change.

J. Emilio Jimenez-Roldan$^{1,2}$ S. A. Wells$^1$ R. B. Freedman$^2$ and R. A. Roemer$^1$

1 Department of Physics and Centre for Scientific Computing, University of Warwick, Coventry, CV4 7AL, UK
2 School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK
E-mail: j.e.jimenez@warwick.ac.uk

Abstract. Our simulations of protein motion are based on a rigidity-based coarse graining criteria which is able to identify flexible and constraint regions and on Normal Modes of Motion (NMM) which map out the directions of motion for a given network. The NMM and rigidity constraints are integrated together by the FRODA module to identify large scale conformational changes at low computational cost. In this context we identify the limits for conformational change of a large multi domain protein -Protein Disulphide Isomerase, by re-assessing the principal NMM at intermediate points along a trajectory as the protein moves over the trajectory defined by the initial principal NMM.

1. Introduction
Although computationally expensive, protein simulations to explore dynamics, fluctuations and motions via molecular dynamics (MD) have proved to be a valuable tool for investigating molecular motions up to a few hundred nanoseconds. However, there is a growing interest in studying processes involving larger proteins and processes of the order of milliseconds or longer to complete. Coarse grained (CG) models are developed with the motivation of simulating large systems for longer time scales at a lower computational cost in order to have a dynamic rapport with experimental observables. Hence, CG models average out atomic degrees of freedom based on individual CG constraints, so that it is possible to investigate large mesoscopic phenomena which are usually beyond the scope of atomistic detailed models. Ideally a good CG model will reproduce the physical behaviour of an experiment or a more computationally costly simulation. Hence, the value of a CG model is defined by how well it captures the essential features of a molecular network and simplifies the handling of the network so that it brings both time and performance advantage. Since CG models allow one to explore larger systems and larger time scales, they have been used most often to study issues on protein folding, conformational change and processes that involve macromolecular interactions involving large time scales.

There is a growing research focus on the role of CG models in interpreting experimental data [1], protein rigidity [2, 3], protein folding [4], multi-scale methods [5], and membrane proteins [6]. In particular there has been a development of CG models which specialise in protein flexibility and in exploring their accessible conformational space [7, 8]. Such methods
have been used to simulate conformational changes either by using stereochemically acceptable pathways [9] or normal modes of motion [10–13]. However, there is no method that will identify the limits of the accessible conformational space for proteins. In this paper we present the preliminary results of our CG simulation strategy for the principal Normal Mode (pNM) which is defined as the lowest frequency non-trivial NMM. We investigate the limits of Protein Disulphide Isomerase (PDI) conformational change for its principal mode while using our integrated coarse grained method. Further, we investigate the distance that PDI’s domains can move freely without steric clashes or network bond constraints along the trajectory defined by the pNM and finally we show that it is possible to to study large conformational changes at a computational cost of minutes using our CG strategy.

Protein Disulphide Isomerase is a 522 amino acid protein with four domains a-b-b’-a’ plus a linker region between b’ and a’ and a C-terminal extension by a domain [14]. PDI catalytic activity is essential in the formation of disulphide bonds in secretory proteins and it is required to be a conformationally flexible molecule to perform its catalytic activity [15].

2. Methods

Figure 2. PDI’s RCD graph defined for its four domains a-b-b’-a’. Within each domain, a colour defines a rigid cluster. The energy cut off increases downwards, so each horizontal line is created when there is a change in the protein rigidity as we remove the weakest hydrogen bond and test the network rigidity. The rigid residues are defined by a colour thick line and the flexible ones by black thin line.
We obtain the Rigid Cluster Decomposition (RCD) shown in Figure 1 using FIRST [7], a software that identifies rigid and flexible regions or clusters in a protein for a given energy cut off. Such clusters are used to coarse grain the simulation and are considered to move as a single unit. The set of NMM are obtained using ElNemo software [16,17]. The directed motion for PDI is simulated using the software FRODA [18] which integrates the NMM and the rigidity constraints from the RCD to provide a CG simulation at a very low computational cost. As the protein structure moves along the conformational path defined by a given initial NMM the new conformational structure is recorded at given intervals or frames.

Each frame represents a computer cycle in which the protein is exploring the conformational space in the direction given by a NMM. We record a new structure for every 100 frames and up to 2500 frames. For each of these structures we calculate the new set of NMM particular to that new conformation which I refer to as intermediate NMM. This process is illustrated in Figure 2 where a set of three structures show PDI closing (from left to the right) and the intermediate NMM (M_i-M_i-‘M_i”) calculated for each frame. The positive and negative direction of motions are chosen arbitrarily.

Thereafter, we calculate how each intermediate set of NMM have changed with respect the initial NMMs by obtaining the dot product between the vector that defines each of the intermediate NMM and the initial NMM. Even though in this paper we present the assessment of the pNM only, we show how it is possible to identify changes between the initial pNM and the intermediate pNM calculated as the protein moves along the trajectory defined by the initial pNM.

3. Results
The dot product between the initial pNM and the intermediates pNM shown in Figure 3-Left reveal that there is a number of frames for which the initial pNM holds good correlation with the intermediates pNM. This is true for all the energy cut off selected. Hence we can identify the range of conformational change that the protein is free from steric effects and for which it moves freely over the initial pNM. Using the intermediate structural conformations obtained

![Figure 3. (a) Dot product between the pNM calculated initially and the pNM calculated at each of the 2500 frames, for the positive and negative direction of motion. We include calculations for six different energy cut off each of which include a different number of bonds and therefore a different rigidity patter. (b) Calculated distance between the two active sites, within the a and a’ domain versus frames over the two directions of motion for the pNM. The minimum distance remains similar for different energy cut off but the maximum distance increases for higher energy cut off.](image-url)
using the pNM we show in Figure 3-Right the variation in distance between the two PDI active sites in domains $\alpha$ and $\alpha'$ as the motion over the initial pNM progresses over each frame. The 3D representation of PDI’s conformational change and its active sites relative motion is show in Figure 4. The simulated conformational change of PDI over the pNM trajectory reveals that the distance between PDI’s active sites spans from approximately $14\AA$ to over $55\AA$ depending on the energy cut off, Figure 3-Right. The difference in the accessible conformational space between different energy cut off may be due to the large number of hydrogen bonds which are present at low energy cut off which would make the bond network less flexible to the point of restricting protein mobility for a lower energy cut off.

4. Discussion
The simulation of conformational changes for large proteins are out of reach for most atomistic models. Our coarse grained simulations however are able to map out PDI’s conformational change along the pNM trajectory but also, equally important, the simulation time scale is of a few hours. Furthermore, the results here presented are in agreement with previous experimental work stating that PDI is required to be a conformationally flexible molecule to perform its catalytic activity [15].

Acknowledgements
We gratefully acknowledge support by EPSRC and BBSRC.

References
[1] Russel D, Lasker K, Phillips J Schneidman-Duhovny D, Velazquez-Muriel J A and Sali A 2009 Curr Opin Cell Biol, 2197-108
[2] Wells S A, Jimenez-Roldan J E, Roemer R A 2009 Phys. Biol. 6 046005
[3] Clementi C 2008 Curr Opin Struct Biol 2008 18 10-15
[4] Shervood P, Brooks B R and Sansom M S P 2008 Curr Opin Struct Biol 18 630-640
[5] Yaliraki S N and Barahona M 2007 Philos Trans R Soc A Math Phys Eng Sci 365 2921-2934
[6] Bahar I, Lezon T R, Bakan A and Shrivastava I H 2010 Chem Rev 110 1463-1497
[7] Jacobs D J, Rader A J, Thorpe M F and Kuhn L A 2001 Proteins: Struct Funct Genet 44 150-165
[8] Raveh B, Enosh A, Schueler-Furman O and Halperin D 2009 PLoS Comput Biol 5 e1000295
[9] Farbeel D W, Speranski K and Thorpe M F 2010 Proteins 78 2908-2921
[10] Moritsugu K, Njunda B M and Smith J C 2010 J Phys Chem B 114 1479-1485
[11] Rueda M, Chacn P and Orozco M 2007 Structure 15 565-575
[12] Nakasako M, Maeno A, Kurimoto E, Harada T, Yamaguchi Y, Oka T, Takayama Y, Iwata and Kato K 2010 Biochemistry 49 6953-6962
[13] Dykeman E C and Sankey O F 2010 J. Phys. Condens. Matter 22 423202
[14] Freedman RB, Klappa P and Ruddock LW 2002 EMBO reports 3 2 136-140
[15] Tian G, Kober F, Lewandrowski U, Sickmann A, Lennarz W J and Schindelin H 2008 The J. of Biological Chemistry. 283 33630-33640
[16] Suhre K and Sanejouand Y H 2004 Nucleic Acids Research 32 W610-W614
[17] Suhre K and Sanejouand Y H 2004 Acta Cryst D 60 796-799
[18] Wells S, Menor S, Hespenheide B and Thorpe M F 2005 Phys. Biol. 2 S127