TREATMENT OF DISULFIDE BONDS IN COARSE-GRAINED UNRES FORCE FIELD

PAWEŁ KRUPA

Faculty of Chemistry, University of Gdansk
Wita Stwosza 63, 80-952 Gdansk, Poland

Institute of Physics, Polish Academy of Sciences
Aleja Lotnikow 32/46, 02-668 Warsaw, Poland

(received: 12 September 2016; revised: 7 October 2016;
accepted: 14 October 2016; published online: 28 October 2016)

Abstract: Disulfide bonds, despite the advances of the computational methods, are underrepresented in theoretical chemistry and the role of disulfide bonds is often diminished in bioinformatical studies. Most of the molecular modeling tools do not allow studying the process of disulfide bond formation and breaking, which is equally important as the sole presence of disulfide bonds in proteins and peptides. The UNRES (UNited RESidue) coarse-grained force field allows treating disulfide bonds in two ways: as static (formed or broken in the simulation) or dynamic (all specified cysteine residues can form and break disulfide bonds during simulation). The comparison between those two approaches of disulfide-bond treatment is presented for protein folding on the example of four small β- and $\alpha+\beta$ proteins with one, two, three and four disulfide bonds. The results clearly show that proper disulfide bond treatment is important in simulations and significantly enhances the quality of folded structures.

Keywords: molecular dynamics, protein folding, disulfide bonds, coarse-grained force field
DOI: https://doi.org/10.17466/tq2016/20.4/l

1. Introduction

Disulfide bonds and other cross-links [1] play important role in protein folding [2], protein structure stabilization [3] and often are responsible for the function of proteins and peptides. [4] Although over 23% of the proteins stored in the PDB database [5] contains at least one disulfide bond, their presence in bioinformatical studies is often omitted due to technical difficulties. In most of the existing force fields disulfide bonds can be present (usually in form of a restraint function clipping two sulfur atoms at the desired distance) or cannot be present for the whole time of the simulation and their formation or reduction cannot be studied. UNited RESidue (UNRES) [6–9] is a coarse-grained model of proteins,

a part of the unified coarse-grained model for biological macromolecules [9], which allows treating disulfide bonds either as a simple restraint present for the whole time of the simulation ('static disulfide bonds') or as bimodal potential allowing disulfide bonds to be formed and broken during the simulation ('dynamic disulfide bonds'). [10] Dynamic treatment of disulfide bonds allows us to study the protein folding process with subsequential oxidization and reduction of native and non-native disulfide bonds or investigate the stability of disulfide bonds in a reductive environment. The influence of the different treatments of disulfide bonds on the folding of four small β and $\alpha+\beta$ proteins: 2JNI, 5JHI, 2LXZ, and 2M8B containing one, two, three and four disulfide bonds, respectively (Figure 1), was assessed.

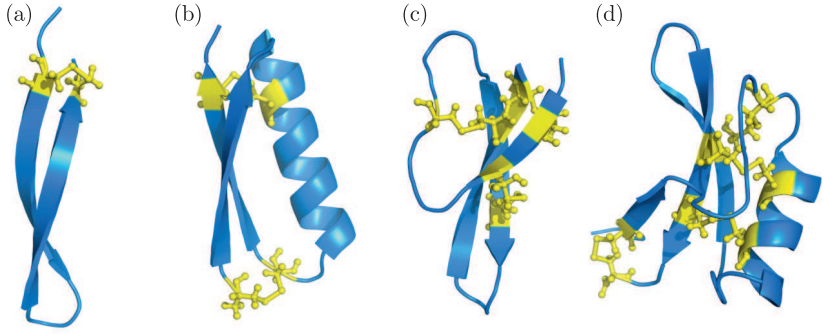

Figure 1. Cartoon representations of: (a) 2JNI and (b) 5JHI, (c) 2LXZ, (d) 2M8B with yellow ball-and-sticks indicating cysteine residues forming disulfide bonds

2. Methods

UNRES (UNited RESidue) [6–9] is a coarse-grained model for proteins which simplifies the polypeptide chain into two interaction sites per residue: spherical united peptide groups (p's) located halfway between two consecutive Cα atoms (which are not interaction sites and define the geometry only) and united side chains (SC's) (Figure 2). Due to a reduced number of interaction sites and averaged secondary degrees of freedom, the UNRES force field provides at least a 3 order of magnitude speed-up comparing to the all-atom simulations in water [11]. Static disulfide bonds were treated as flat-bottom restraints between SCs of the respective cysteine residues in the range of 4 to 5 Å from each other. Dynamic disulfide bonds were treated as in ref [10] with an additional energy barrier to prevent forming triple-sulfide bonds [12]. The newest version of the UNRES force field with recently implemented local potentials coupling orientations of side-chains with the backbone [13, 14] was used, which was reoptimized on seven training proteins [12], using the maximum likelihood method [15].

Series of the Multiplexed Replica Exchange Molecular Dynamics (MREMD) [16, 17] simulations were performed to study the influence of disulfide models on the folding and stability of protein structures. MREMD is an extension of the

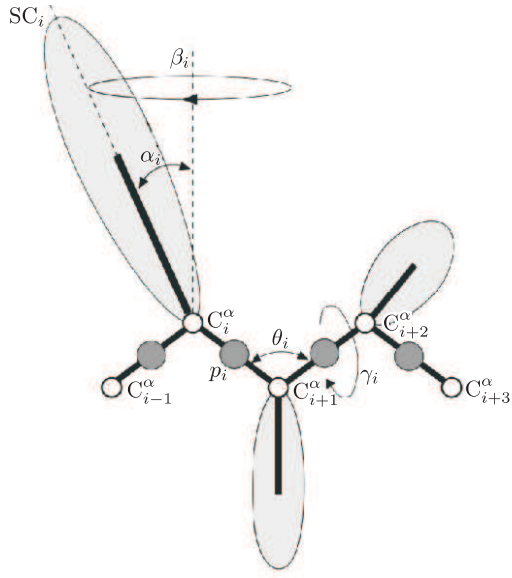

Figure 2. Schematic representation of the UNRES model of a polypeptide chain in which there are two interaction sites per residue: a united side-chain (SC) and a united peptide group (p) represented by light-grey ellipses and dark-grey circles, respectively. C^α atoms (white circles) and angles β, α, Θ and γ define the positions of backbone and side-chains

Replica Exchange method [18, 19] in which more than one trajectory is run in given temperature resulting in improvement of the conformational search and scalability of the simulations. [17] Each MREMD simulation started from a fully extended polypeptide chain and consisted of 60 trajectories, two per temperature: 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 295, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 480, and 500 K. Each trajectory was calculated for 50 000 000 steps, each was equal to 4.89 fs. Every 10 000 steps exchanges of the replica were attempted and snapshots were saved for further analysis. Structures corresponding to the last 40 000 000 steps were used in the Weighted Histogram Analysis Method (WHAM) [20] to derive the temperature-depended ensembles, which were subsequently clustered into five groups using the minimum-variance clustering method in a temperature below the heat capacity peak, as in a previous work [21]. The root-mean square deviation (RMSD) of the (i) best snapshot ($\mathrm{RMSD_{min}}$), (ii) highest RMSD of thousand best snapshots ($\mathrm{RMSD_{1000}}$) and the average cluster structure RMSD ($\mathrm{RMSD_{clust}}$) were used for the analysis, as in a previous work [22].

3. Results and discussion

Analysis of MREMD simulations in a UNRES force field for four tested proteins shows that the presence of disulfide bonds improves the quality of the resulting structures reflected in a decrease in all the calculated RMSD values (Figure 3). Disulfide bonds decrease the flexibility of the chain by restricting

the distances between cysteine side-chains in simulations with the use of static disulfide bonds and by introducing a small but important energy barrier necessary to break the disulfide bond in simulations with the use of dynamic disulfide bonds. It can be also observed by greater differences in the quality between average structures (RMSD_{1000} and $\mathrm{RMSD}_{\mathrm{clust}}$) and the best structures observed in the simulations ($\mathrm{RMSD}_{\mathrm{min}}$). The $\mathrm{RMSD}_{\mathrm{clust}}$ values are usually better in simulations with dynamic treatment of disulfide bonds, especially in comparison to RMSD_{1000}, due to the applied WHAM procedure, which favors lower-energy structures in clustering.

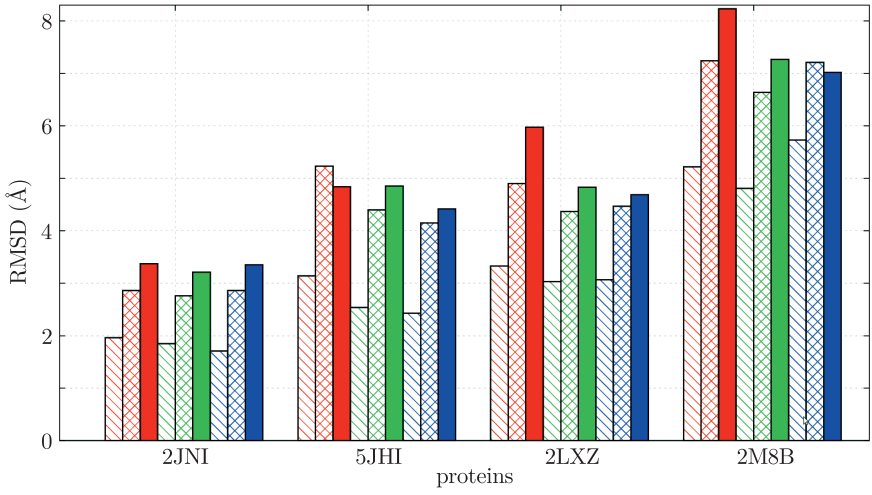

Figure 3. RMSD values for comparison of resulting structures without disulfide bonds (red), with static disulfide bonds (green) and with dynamic disulfide bonds (blue) with the experimental structures; best structures ($\mathrm{RMSD}_{\mathrm{min}}$) are represented by single slanted pattern bars, 1000 lowest rmsd structures (RMSD_{1000}) as slanted-checker bars and cluster average structures ($\mathrm{RMSD}_{\mathrm{clust}}$) as solid bars

It can be observed that with the increased number of Cys residues the performance of the dynamic treatment of disulfide bonds starts to suffer due to a rapid increase in possible combinations, equal to $(2n-1)!!$, which $2n$ Cys residues might probe while forming n disulfide bonds [23]. In the case of static disulfide bonds low performance for 2M8B likely results from the fact that some trajectories get stuck in misfolded high-energy conformations and cannot overcome energy barriers caused by the disulfide-bond restraints.

4. Conclusions

For every tested protein, the presence of disulfide bonds increases the foldability of the proteins and accelerates the calculations by restricting the conformational space needed to be searched. While static treatment of disulfide bonds provides unphysical folding pathways by causing rapid collapse of the molecules due to the high restraint-penalty values, it also provides the best results

by enforcing only native disulfide bonds. It is also easy to implement and use. Dynamic treatment of disulfide bonds allows forming non-native disulfide bonds that are observed experimentally [2], which slows down the folding process in comparison to the static disulfide bonds, however, it allows studying undisturbed folding pathways. Therefore, while even static disulfide bond treatment improves the quality of predicted structures, it should be used only for limited purposes, *e.g.* when it is the final structure and not the folding pathway that is important or when the protein structure is almost rigid – in all other studies dynamic treatment should be used instead.

Acknowledgements

I thank Adam Liwo Ph. D. and Adam Sieradzan Ph. D. for helpful discussions.

References

[1] Mozolewska M A, Sieradzan A K, Niadzvedstki A *et al.* *J Biomol Struct Dyn.* doi: 10.1080/07391102.2016.1234414
[2] Qin M, Wang W, Thirumalai D 2015 *Proc Natl Acad Sci USA* **112** 11241 doi: 10.1073/pnas.1503909112
[3] Betz S F 1993 *Protein Sci* **2** 1551 doi: 10.1002/pro.5560021002
[4] Hogg P J 2003 *Trends Biochem Sci* **28** 210 doi: 10.1016/S0968-0004(03)00057-4
[5] Berman H M 2000 *Nucleic Acids Res* **28** 235 doi: 10.1093/nar/28.1.235
[6] Liwo A, Pincus M R, Wawak R J *et al.* 1993 *Protein Sci* **2** 1715 doi: 10.1002/pro.5560021016
[7] Liwo A, Ołdziej S, Pincus M R *et al.* 1997 *J Comput Chem* **18** 849 doi: 10.1002/(SICI)1096-987X(199705)18:7<849::AID-JCC1>3.0.CO;2-R
[8] Liwo A, Czaplewski C, Ołdziej S *et al.* 2008 *Simulation of protein structure and dynamics with the coarse-grained UNRES force field*, *Coarse-Graining Condens Phase Biomol Syst* 107
[9] Liwo A, Baranowski M, Czaplewski C *et al.* 2014 *J Mol Model* **20** 2306 doi: 10.1007/s00894-014-2306-5
[10] Chinchio M, Czaplewski C, Liwo A *et al.* 2007 *J Chem Theory Comput* **3** 1236 doi: 10.1021/ct7000842
[11] Khalili M, Liwo A, Rakowski F *et al.* 2005 *J Phys Chem B* **109** 13785 doi: 10.1021/jp058008o
[12] Krupa P *et al.* (in preparation)
[13] Krupa P, Sieradzan A K, Rackovsky S *et al.* 2013 *J Chem Theory Comput* **9** 4620 doi: 10.1021/ct4004977
[14] Sieradzan A K, Krupa P, Scheraga H A *et al.* 2015 *J Chem Theory Comput* **11** 817 doi: 10.1021/ct500736a
[15] Zaborowski B, Jagieła D, Czaplewski C *et al.* 2015 *J Chem Inf Model* **55** 2050 doi: 10.1021/acs.jcim.5b00395
[16] Liwo A, Khalili M, Czaplewski C *et al.* 2007 *J Phys Chem B* **111** 260 doi: 10.1021/jp065380a
[17] Czaplewski C, Kalinowski S, Liwo A, Scheraga H A 2009 *J Chem Theory Comput* **5** 627 doi: 10.1021/ct800397z
[18] Sugita Y, Okamoto Y 1999 *Chem Phys Lett* **314** 141 doi: 10.1016/S0009-2614(99)01123-9
[19] Nanias M, Czaplewski C, Scheraga H A 2006 *J Chem Theory Comput* **2** 513 doi: 10.1021/ct050253o

[20] Gallicchio E, Andrec M, Felts A K, Levy R M 2005 *J Phys Chem B* **109** 6722 doi: 10.1021/jp045294f
[21] Krupa P, Mozolewska M A, Wiśniewska M *et al.* 2016 *Bioinformatics* **32** 3270 doi: 10.1093/bioinformatics/btw404
[22] Krupa P, Mozolewska M A, Joo K *et al.* 2015 *J Chem Inf Model* **55** 1271 doi: 10.1021/acs.jcim.5b00117
[23] Rubinstein R, Fiser A 2008 *Bioinformatics* **24** 498 doi: 10.1093/bioinformatics/btm637