Published in final edited form as:

J Struct Funct Genomics. 2010 June ; 11(2): 101–111. doi:10.1007/s10969-010-9079-6.

Enhanced crystal packing due to solvent reorganization through reductive methylation of lysine residues in oxidoreductase from *Streptococcus pneumoniae*

Yao Fan and Andrzej Joachimiak

Biosciences Division, Midwest Center for Structural Genomics and Structural Biology Center, Argonne National Laboratory, 9700 S Cass Ave., Argonne, IL 60439, USA

Andrzej Joachimiak: andrzejj@anl.gov

Abstract

Protein crystallization is in part driven by the changes in the entropy of the system, but opinions differ as to whether the solute (protein) or solvent (water) molecules make more of a contribution to the overall entropic change. Methylation of lysine residues in proteins has been used to enhance protein crystallization. We investigated using molecular dynamics simulations with explicit solvent molecules, the behavior of several native proteins and their methylated counterparts chosen from an earlier large-scale study. Methylated lysines are capable of making a variety of interactions including H-bonds with protein residues and solvent. We demonstrate that methylation on the lysine slightly increases its side chain conformational entropy by about $3.5 \text{ J mol}^{-1} \text{ K}^{-1}$. Analysis of the radial and spatial distributions of the water molecules around the methylated lysine surface in oxidoreductase from *Streptococcus pneumoniae* revealed a larger sphere of water molecules with low entropy, as compared with solvent associated with unmethylated lysine. If methylated lysine were to make interactions at the protein–protein interface, the low-entropy water molecules associated with methylated lysines would be released, resulting in a gain of entropy. We show that this gain more than compensates for the loss of protein entropy. Therefore, we propose that lysine methylation favors the formation of crystals through solvent entropic gain.

Keywords

Reductive methylation; Methylated lysine; Protein crystallization; Protein interactions; Protein modification

Introduction

Protein crystallization has long been the major bottleneck in structure determination using single crystal X-ray crystallography, and it remains as such in spite of many advances in the structural genomics era [1]. Typically, in order to obtain a good quality single crystal, protein is screened against hundreds to thousands of conditions with varying buffers, salt, ionic strength, pH, precipitant, temperature, ligands, various additives, etc. This approach is

© US Government 2010

Correspondence to: Andrzej Joachimiak, andrzejj@anl.gov.

The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

currently the standard procedure in almost all structural biology labs, but its limitation has received increasing attention from some research groups [2,3]. For example, it was suggested that this approach may help to select for proteins that have high intrinsic propensity to crystallize [2]. For the majority of proteins that fail “condition screening,” changing the buffer might never lead to crystals.

In an alternative approach, instead of changing crystallization conditions one can change the properties of the protein itself to enhance its crystallizability. Several approaches have been employed for re-engineering the target protein (partial proteolysis, both in vitro and in situ, site-directed mutagenesis, chemical modification, the use of sequence orthologs) with varying degrees of success [4–6]. Chemical modification of proteins using reductive methylation has the advantage that it is very simple and it does not require additional protein expression and purification. Recently a large-scale investigation was carried out at the Midwest Center for Structural Genomics (MCSG) to assess the effect of lysine methylation. 370 unique proteins that failed to crystallize were chosen for reductive methylation treatment, and out of which 26 crystal structures of methylated proteins were determined. The method has ~7% success rate, but often produces crystals that diffract to high resolution, resulting in better structures [7]. Despite its success very little is known about the underlying mechanism that contributes to the enhanced crystallization.

Our ability to predict the crystallization behavior of the protein molecules is largely plagued by our limited understanding of the crystallization process itself and by the horrendous task of predicting crystal packing contacts from protein sequence. The overall theoretical framework of the events involved in protein crystallization has been established [8,9]. The free energy change of the crystallization process can be written as:

$$\Delta G_{\text{crystallize}} = \Delta H_{\text{crystallize}} - T\Delta S_{\text{crystallize}} \quad (1)$$

For most proteins, the enthalpy change, $\Delta H_{\text{crystallize}}$, is near zero or positive, which leaves the entropy change, $\Delta S_{\text{crystallize}}$, as the predominant driving force for crystallization. The overall entropy change can be further decomposed into the contributions from solute (protein) and solvent (water), respectively:

$$\begin{aligned} \Delta S_{\text{crystallize}} &= \Delta S_{\text{protein}} + \Delta S_{\text{water}} \\ &= (S_{\text{protein}}^{\text{crystal}} - S_{\text{protein}}^{\text{aq}}) + (S_{\text{water}}^{\text{free}} - S_{\text{water}}^{\text{localized}}) \end{aligned} \quad (2)$$

The first term $\Delta S_{\text{protein}}$ is the entropy *loss* of the protein molecule in the transition from aqueous solution to the crystalline state, which originates from the decrease of the degrees of freedom of the backbone and side chain atoms of the protein. The second term ΔS_{water} , on the other hand, represents the entropy *gain* due to the mobilization of protein-bound water molecules from the hydration shells surrounding the protein molecule into the bulk solvent during crystallization. In principle, based on this framework, there are two main strategies to engineer the target protein to enhance its propensity to crystallize: (i) minimize the entropy loss by decreasing the entropy of protein molecule in the aqueous solution, (ii) maximize the entropy gain by increasing the number of water molecules that will be released and/or enhancing their localizability. It has been shown in previous studies that mutating lysine (also Glu and Gln) residues on the surface of the protein molecule into alanine can yield diffraction quality crystals [5,10]. This phenomenon was attributed to the fact the Lys → Ala mutation can reduce the side chain entropy of the protein and facilitate better crystal packing near the site of mutation [10–12]. Since reductive methylation involves the lysine residues that are solvent exposed, it

is intriguing to find whether it may use the same mechanism as proposed for the Lys $\rightarrow$ Ala mutation.

Interestingly, methylation of lysine residues plays also an important cellular function in chromatin remodeling, signaling and transcription regulation and is implicated in several diseases. The properties of methylated lysine and its interaction with proteins and solvent may have important implication to better understanding these processes [13,14].

In the present study, the effects of reductive methylation on the protein molecule and the surrounding solvent molecules are investigated using molecular dynamics (MD) simulation on both the methylated and unmethylated versions of a protein molecule, the structures of which were determined at MCSG. Before the simulations could be carried out, the molecular mechanics force field parameters for the dimethyl lysine (dmLys) were optimized to replicate the results from ab initio quantum mechanical calculations. From the MD trajectories, the properties of the protein molecules, including its global dynamics and the local motions near the methylation site, can be evaluated. More detailed analyses were carried out on the methylation site, so the values of the side chain conformational entropy of both methylated and unmethylated lysine residues can be calculated. Finally, we attempted to elucidate how addition of two methyl groups to lysine may impact the properties of solvent molecules near the site chain, and how it can weigh on the protein propensity for crystallization.

Materials and methods

Bi-molecular complex interaction energies

The interaction energies between the solute molecule (ethyl amine as a model for lysine and N,N-dimethylethyl amine for dmLys) and the solvent molecule (water) were estimated with the procedure similar to those employed in the development of molecular mechanics force field parameters. The geometry of both the solute and solvent molecules were optimized at the Hartree–Fock 6–31G(d) level of theory, and the hydrogen bonding complexes were constructed as shown in Fig. 1. In all three cases, the oxygen atom in the water molecule serves as the hydrogen bond acceptor and the only geometrical parameter being optimized in the solute–solvent complex is the distance between the hydrogen bond donor (nitrogen for the first complex and carbon for the other two) and acceptor (oxygen from the water molecule). All calculations were carried using the GAMESS [15] package.

Parameterization of dmLys

The parameters of the native lysine are included in the standard CHARMM [16] package, and the parameters of dmLys were prepared following the established procedure for optimizing force field parameters as detailed in previous works [17,18]. The optimized parameters for tetraethylammonium cation developed by Roux and co-workers [19] were used as the references for the bonded energy terms. The non-bonded energy terms, including the atomic partial charges and the van der Waals parameters, were adjusted to reproduce the interaction energy values calculated at the Hartree–Fock 6–31G(d) level.

Molecular dynamics simulation of proteins

Model molecule—The Gfo/Idh/MocA family oxidoreductase from *Streptococcus pneumoniae* was chosen as the target molecule for simulation and analysis. This protein was selected because the high-resolution crystal structures of both the unmethylated and methylated forms are available (PDB ids 2HO3 and 2HO5, respectively). Also, the residue Lys138 is unambiguously doubly methylated (dmLys) and is located near the subunit interface in both crystal forms, highlighting its importance in crystal packing. The two crystal structures have missing residues in the loop region between residues 154 and 171. These missing residues were

modeled using the loop-building module in COOT to fit into the residual electron density. Hydrogen atoms were added to the heavy atoms using the HBUILD module in CHARMM. The protonation states of the side chains of titratable residues were set as follows: the side chains of Lys and Arg are fully protonated, Asp and Glu side chains are fully deprotonated, and the side chain of His is set to be neutral. The side chain of the methylated lysine residue, which shows a pKa range between 9.29 and 10.23 [20], is also set as protonated on its ϵ -nitrogen atom. The all-atom CHARMM27 force field [18] parameters were used for the protein molecules.

The protein model with the hydrogen atoms added were then submerged in a 120 Å by 120 Å by 120 Å cubic box of pre-equilibrated TIP3P [21] water molecules. The center of mass of the protein molecule is placed at the center of the water box, and the longest axis aligned along the x-axis. All water molecules whose oxygen atoms are less than 2.5 Å away from heavy atoms of the protein molecule are deleted. A total number of 17 sodium ions were added to the water box to counter the negative charge of the protein molecule. A 9 Å cutoff was used for the van der Waals non-bond interactions, and the particle-mesh Ewald (PME) method [22] was used to calculate the electrostatic interactions. All bonds involving hydrogen atoms were constrained using SHAKE [23].

The fully solvated system was heated to 300 K with an increment of 50 K for every 500 ps. A harmonic restraint was placed on the heavy atoms of the protein, which was gradually released along the course of heating. After the system reached the desired temperature, it was subject to a further 3 ns of equilibration at constant temperature (300K) and pressure (1 atm). Once the system has reached equilibrium, as judged from criteria such as temperature, energy, and water box volume, trajectory data were collected from the subsequent 13 ns' simulation. All simulations were carried out using the NAMD [24] package.

Side chain conformational entropy—The side chain conformational entropy of the doubly methylated and unmethylated lysine side chains were calculated using the method proposed by Sternberg and coworkers [25,26]. As shown in Fig. 2, the torsional entropy of a C–C single bond is evaluated as

$$S^{torsion} = -R(p_{gauche+} \ln p_{gauche+} + p_{trans} \ln p_{trans} + p_{gauche-} \ln p_{gauche-}) \quad (3)$$

where p denotes the probability of the bond being in the *gauche+*, *trans* and *gauche-* states, respectively. For an amino acid side chain with multiple C–C single bonds, the overall conformational entropy is the summation of the individual bond torsional entropies:

$$S^{conf} = \sum_{i=1}^n S_n^{torsional} \quad (4)$$

For Lys and dmLys residues as shown in Fig. 3 with torsional angles χ , their side chain conformational entropy can be expressed as

$$S^{conf} = \sum_{torsions} S_{torsions} = \sum_n \left(-R \sum_{i=1}^3 p_i(\chi_n) \ln p_i(\chi_n) \right) \quad (5)$$

Water density and entropy distribution—To calculate the density of water molecules in the vicinity of the methylation sites, the following method was employed. First, the space

around the methylated lysine residue is divided into a series of concentric spherical shells with 1 Å thickness, centered on the ε-nitrogen of the lysine residue. Then, for each saved structure in the simulation trajectory, the number of water molecules are counted for every shell, and averaged over the simulation time.

The entropy of water molecules in the hydration shell around the protein molecule was estimated with a grid-based direct counting method [27]. The space around the solute molecule is divided into 1 Å cubic grids, and each grid point is associated with a finer mesh with a dimension of 0.2 Å, which takes into account the distribution of water molecules around the grid point. For each grid point, the relevant water entropy can be expressed as the following equation:

$$S = -R \sum_{l,m,n} P_{l,m,n} \ln P_{l,m,n} \quad (6)$$

where $P_{l,m,n}$ is the probability of finding a water molecule in the finer mesh point l, m, n surrounding the grid point, and R is the gas constant. The entropy is then plotted as a function of distance from the center atom, which is the ε-nitrogen of Lys (or dmLys) residue.

Results

DmLys-water interaction energy

The interaction energy of the bimolecular complex formed by a solute (represented by ethyl ammonium, dimethyl ethyl ammonium and dimethyl ethyl ammonia) and a solvent (TIP3P water) molecule is a good indicator of the length and strength of the intermolecular hydrogen bonds, if they form. As shown in Fig. 1, the nitrogen atom in the ethyl ammonium cation forms a strong hydrogen bond by donating a hydrogen atom to the oxygen of the water molecule. The interaction energy reaches its minimum of $-18.4 \text{ kcal mol}^{-1}$ when the distance between the hydrogen donor (nitrogen in ammonium) and acceptor (oxygen in water) is 2.84 Å. For the unprotonated dimethyl ethyl ammonia, the interaction energy with water is almost negligible. When the ammonia is protonated, however, the interaction energy profile changes substantially, with a minimum of $-9.6 \text{ kcal mol}^{-1}$ when the distance between the carbon and oxygen atoms is 3.28 Å. This implies that, when the nitrogen atom in dimethyl ethyl ammonia is protonated, the two methyl groups attached to it are polarized and can interact with polar atoms such as oxygen and nitrogen. Consequently, the interactions between the methyl groups and the water molecule exhibit certain traits of hydrogen bonds, albeit weaker in strength. The existence of such 'quasi' hydrogen bonds around the methylated lysine side chain is confirmed by the available crystallographic data, where the distances between the η-carbon and nearby hydrogen bond acceptors are consistently between 3.1 and 3.4 Å [7,28].

Local hydrogen bond analysis near the methylation site

There are two aspartate residues, Asp137 and Asp242, near the methylation site, Lys138, in the model protein (Fig. 4), and the unmethylated and methylated lysine residues exhibit different interaction patterns with them. Because the two aspartate residues are over 8 Å apart, the unmethylated lysine side chain can form a hydrogen bond with one or the other, but not with both; on the other hand, the methylated lysine, with a radius of interaction that is about 1.4 Å longer, can potentially interact with both Asp137 and Asp242. This is evidenced by monitoring the distance between the NE atom of Lys/dmLys138 and the carboxylate oxygen atoms (OD1 and OD2) of Asp137 and Asp242 throughout the simulation.

As shown in Fig. 5a and b, in the unmethylated protein, at the beginning of the simulation, a hydrogen bond exists between Lys138 and Asp242 (Fig. 4a), which is broken after 9.5 ns into the simulation (Fig. 4b). The breaking of the hydrogen bond between Lys138 and Asp242, however, is accompanied by the simultaneous formation of another hydrogen bond, between Lys138 and Asp137 (Fig. 4b). Due to its flexibility, the side chain of Lys138 can rotate between different conformations and alternate between mutually exclusive hydrogen bonds.

For the methylated protein, the hydrogen bond between dmLys138 and Asp242 remains largely undisturbed during the simulation (Fig. 5a, b), and the distances between the NE of dmLys138 and OD1/OD2 of Asp137 are larger than 3.5 Å (Fig. 5c, d), rendering it impossible to form hydrogen bonds between Asp137 and dmLys138 through the NE atom of the latter. However, the distance between the η -carbon of dmLys138 and OD1 of Asp137 is found to be fluctuating around 3.3 Å, within the optimal range for a methyl group-based hydrogen bond (Fig. 4c) [29–31].

Side chain conformational entropy of the lysine residue

Using Eq. 5, the side chain conformational entropy of Lys and dmLys were calculated from the trajectories of the MD simulations. To rule out the possibility that the results are artifacts due to inappropriate settings in the protein simulation, entropy values were calculated from the simulations of free Lys and dmLys using the same method. The calculated entropy value of free Lys was comparable with values from early experimentally and theoretically determined values [25,26], thereby validating the method being employed.

As shown in Table 1, from the simulation with the whole protein molecule, dmLys shows a higher side entropy value than the unmethylated counterpart by about $3.5 \text{ J mol}^{-1} \text{ K}^{-1}$. When the protein environment is not present, both Lys and dmLys have much larger degrees of freedom and hence larger side chain conformational entropy. In terms of relative entropy value, however, the same trend persists, with dmLys $4.3 \text{ J mol}^{-1} \text{ K}^{-1}$ higher than Lys. It is not unexpected that dmLys has a higher side chain conformational entropy than Lys. In fact, as shown in Fig. 3, after methylation, the threefold symmetry along the CE–NE single bond is lost, and the dmLys residue has one additional torsional angle that must be taken into account, which leads to higher entropy. What is surprising, however, is that in the protein environment, even with the introduction of two bulky methyl groups to the NE atom, dmLys still manages to exhibit more mobility and flexibility than Lys. Table 1 reveals that, the protein environment confines the movement of the Lys and dmLys side chains in the same manner and with almost equal force.

Solvent structure near the methylated lysine

Water is the universal solvent and media for all biological systems. Being encapsulated in a layer of water molecules known as the hydration shell, the protein molecules' behavior is intricately related to the structure and dynamics of the surrounding water molecules [32–34]. For a complex process such as protein crystallization or protein–protein binding, the role being played by the water molecules is crucial and cannot be ignored.

The density of water molecules in a large volume ('bulk solvent') is nearly uniform, and using the dividing scheme described above, the number of water molecules in each spherical shell would be proportional to the volume of the shell, which is in turn proportional to the square of the shell radius. As shown in Fig. 6a, at distances over 6 Å away from the center atom, NE of the lysine residue, the water density distribution is indeed proportional to the square of the distance, but abnormalities arise at shorter distances, where higher density solvent regions disobeying the simple parabolic curve appear. These high-density regions correspond to the

peaks in the water radial distribution function near the ϵ -nitrogen atom of Lys and dmLys (Fig. 6a).

Water entropy distribution

The entropy distribution provides valuable information about the mobility of water molecules, which is not available from the aforementioned water density distribution function. Being in the liquid phase, the water molecules are constantly on the move, but their mobility is different in the solvent shell surrounding the protein molecule than in the bulk solvent. The interactions between the protein surface residues and water molecules act as restraints on the latter, which limit their mobility. As the restraints on a water molecule are reciprocally proportional to its distance to the protein surface, it will exhibit higher mobility—and consequently larger entropy—as it stays farther away from the protein surface. For the unmethylated lysine, only water molecules within a very small distance from the ϵ -nitrogen have lower entropy, and water molecules beyond 3 Å resemble that of the bulk solvent (Fig. 6b). Water molecules near dmLys, however, show a progressive increase of the entropy approaching the bulk solvent, leaving a larger ‘low entropy water zone’ between 3 and 6 Å from the ϵ -nitrogen atom, which coincides with a peak in the water density (Fig. 6b). With lower-than-normal entropy and higher-than-normal density, expelling water molecules in this region to the bulk solvent will incur a larger gain in entropy for dmLys than Lys.

Similar to the calculation of lysine side chain conformational entropy, the water density and entropy distribution around the lysine side chain were calculated for free lysine and dimethylated lysine residues, with nearly identical results as in the model protein system (data not shown). Therefore, we believe that the solvent reorganization through reductive methylation is likely an intrinsic property of the lysine residues itself, which is present regardless of the surrounding protein environment.

Discussion

DmLys is involved in conventional and unconventional hydrogen bonds

As shown in Results, after methylation, the dmLys residue has increased radius of interaction, and its properties changed as well. Similar findings have been reported for histone proteins, where methylation on the key lysine residues and the consequent change of side chain interaction patterns are responsible for protein recognition and binding, as well as chromatin remodeling and gene regulation [35]. These findings naturally lead to the question: is the interaction between the methyl group and nearby oxygen/nitrogen atoms a carbon-based hydrogen bond? The experimental evidence for a carbon-based hydrogen bond from various sources has been reported, ranging from small organic molecules to proteins, and, despite its low strength, it may act as contributing factor to molecular stability and recognition [29,36]. Recent theoretical studies [30,31] have shed more light on the strength and electronic nature of the presumed ‘carbon hydrogen bonds’, which are now sometimes referred to as ‘unconventional hydrogen bonds’ to distinguish from the stronger, ‘conventional’ counterparts.

For the methylated lysine, the distance between the η -carbon and nearby oxygen/nitrogen atoms is found to be between 3.1 and 3.4 Å [7]. Also, the quantum mechanical calculations performed in the present and an earlier study [7] estimate the strength of the interaction to be approximately 9.6 kcal mol⁻¹, much stronger than non-specific van der Waals or Coulomb interactions, yet still weaker than a conventional hydrogen bond. The interactions through the η -methyl groups, therefore, can be categorized as an unconventional hydrogen bond. However, more atomistic details, including the polarization effects on the methyl groups, and the directionality of the hydrogen bond, remain unclear and require more rigorous treatment using

higher levels of quantum mechanical theory, such as the Møller–Plesset perturbation theory [37]. Further work is currently under way.

Entropic contribution to protein binding and crystallization

Derewenda and co-workers have performed extensive site-directed mutagenesis on several recalcitrant proteins, changing lysine (also Glu, Gln) residues to alanine (also to Ser, Thr, Tyr), and managed to obtain high quality crystals of the mutated protein [5,10–12,38]. The crystal structures revealed that the resulting alanine residues are often located along the crystal packing interface, forming head-to-head interactions [10]. The authors concluded that, by mutating lysine to alanine, its side chain conformational entropy is decreased by about $28 \text{ J mol}^{-1} \text{ K}^{-1}$, and there will be less entropy loss as protein molecules pack using this residue [10–12]. However, the Lys $\rightarrow$ Ala mutation amounts to more than just the shrinking of side chain volume, as it also annihilates a positive charge, which may have rather drastic effects given the low dielectric constant in the protein environment. The side chain conformational entropy, therefore, might not be the only underlying factor responsible for the enhanced crystallization behavior of Lys $\rightarrow$ Ala mutants.

In the present study, the methylated lysine residues are also predominantly involved in crystal contact [7], which again raises the question as what is the underlying driving force. Since both dmLys and Lys side chains are positively charged, they will share very similar electrostatic interaction profile. DmLys show higher side chain entropy than Lys by about $3.5 \text{ J mol}^{-1} \text{ K}^{-1}$, which leads to a larger penalty for crystal packing. On the other hand, as shown in Fig. 6, within $6 \text{ \AA}$ from the NE atom of dmLys, there are approximately 10 more water molecules with 20% lower entropy for dmLys than Lys as a result of solvent restructuring. When two dmLys residues form head-to-head contact with each other, water molecules in this region will be expelled to the bulk solvent and gain entropy for the whole system. Since freeing one fully immobilized water molecule into bulk solvent can increase the entropy of the system by $\sim 22 \text{ J mol}^{-1} \text{ K}^{-1}$ [39–41], the entropic gain for dmLys can be estimated to be nearly $4 \text{ J mol}^{-1} \text{ K}^{-1}$ more than Lys for each released water molecule, which is more than compensating for the penalty due to the larger side chain conformational entropy. Therefore, we believe that the reorganization of solvent molecules near the dmLys residue, and the consequent entropy gain from releasing the low-entropy water molecules, instead of the entropy of the protein itself, is the significant factor that renders dmLys suitable for more favorable crystal packing interactions.

Interestingly, a similar mechanism may drive the protein-protein interaction involving methylated lysine with its specific binding sites. For example, the Sex comb on midleg (Scm) protein binds dmLys into a cage that is completely devoided of water molecules, and it uses methyl groups to interact with an ‘aromatic cage’ formed by three conserved nonpolar residues, and the NE atom to make strong hydrogen bond with another conserved Asp residue [42]. The preference of imbedding a methylated lysine side chain over an unmethylated lysine into the binding cage may be attributed to the release of low entropy waters around dmLys.

Methylated lysine as crystal packing ‘hot spots’

Just like small organic or inorganic molecules, the first step in protein crystallization is nucleation, or the alignment and assembly of a few protein molecules. Because protein molecules are macromolecules with a highly anisotropic surface, a stable multi-molecular assembly is only available when the molecules interact through surfaces that can form strong and stable interactions. These surface area patches, known as ‘hot spots’ [43], are scattered throughout the protein surface, and the most stable molecular assembly from nucleation will have the molecules aligned in such a way that the hot spots with the most favorable interaction and minimal steric clashes are at the intermolecular interface.

While the lysine residue is generally not favored in the hot spot, mutating it to alanine or methylating its side chain changes it to the contrary, effectively adding more hot spots to the protein surface and expanding the repertoire of intermolecular interaction patterns. Given the complexity of protein surface shape and interactions, however, the new interaction patterns do not necessarily lead to better crystal packing. As a result, reductive methylation is by no means a 'silver bullet' for all proteins. For some protein molecules, the newly adopted interactions make it possible to crystallize; for some other protein molecules, better crystals become available due to the newly improved crystal packing; for some it may increase aggregation and be detrimental to crystallization, yet for still others, methylation might have no effects on crystallization at all. One possible strategy of manipulating the number and location of hot spots on the protein surface would be to simultaneously mutate some Lys residues and methylate others to create a new pattern of hot spots and at the same time reduce entropy of the protein and increase entropic gain upon formation of new interaction(s).

The protein pair 2HO5 and 2HO3 is a good example of how methylation enhances crystal packing. The loop between residues 154–173 in the protein is highly mobile, being prone to steric clash and disrupting the crystal packing. In the methylated protein, 2HO3, this loop is nestled in the groove formed by two helices, 122–136 and 294–316. This packing mode also places the residue Lys138 at the interface in a head-to-head manner, which will incur undesired electrostatic repulsions. For dmLys, the head-to-head contact can be stabilized by the release of low-entropy water molecules, whereas no stabilizing interactions are present for Lys. As a result, the crystal packing pattern as shown in Fig. 7b is only feasible for the methylated protein, and the unmethylated protein will instead pack in a manner that is sub-optimal for positioning the loop region, with the lysine residues being farther apart (Fig. 7a). The higher quality of the crystal formed by the methylated protein is evidenced by the higher resolution of diffraction data (2.56 vs. 2.0 Å) and lower overall temperature factor (76.24 vs. 7.09 Å²) of the crystal structure.

Conclusion

We have provided a theoretical explanation on the enhanced crystallization behavior of oxidoreductase from *S. pneumoniae* after reductive methylation on lysine residues. Our molecular dynamics simulations demonstrate that the side chain conformational entropy of dmLys is larger than unmethylated lysine, which poses a larger penalty for crystallization. However, this penalty is more than compensated by the entropic gain from the release of water molecules due to the solvent reorganization near the lysine residue after methylation. Contrary to the native lysine, the dimethyl lysine residue becomes a hot spot for inter-molecular interactions, adding to the possible crystal packing patterns, which may lead to the crystallization of intractable protein targets and/or improved crystal quality.

Acknowledgments

We thank members of the Structural Biology Center and Midwest Center for Structural Genomics at Argonne National Laboratory for providing data, and also Argonne Leadership Computing Facility for providing the computing resources. We would like to thank Lindsey Butler for help in preparation of the manuscript for publication. This work was supported by National Institutes of Health Grant GM074942 and by the U.S. Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357.

References

1. Stevens RC. High-throughput protein crystallization. *Curr Opin Struct Biol* 2000;10(5):558–563. [PubMed: 11042454]
2. Jancarik J, Kim SH. Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Crystallogr* 1991;24:409–411.

3. Segelke BW. Efficiency analysis of sampling protocols used in protein crystallization screening. *J Cryst Growth* 2001;232(1–4):553–562.
4. Dong A, Xu X, Edwards AM, Chang C, Chruszcz M, Cuff M, Cymborowski M, Di Leo R, Egorova O, Evdokimova E, Filippova E, Gu J, Guthrie J, Ignatchenko A, Joachimiak A, Klostermann N, Kim Y, Korniyenko Y, Minor W, Que Q, Savchenko A, Skarina T, Tan K, Yakunin A, Yee A, Yim V, Zhang R, Zheng H, Akutsu M, Arrowsmith C, Avvakumov GV, Bochkarev A, Dahlgren LG, Dhe-Paganon S, Dimov S, Dombrowski L, Finerty P Jr, Flodin S, Flores A, Graslund S, Hammerstrom M, Herman MD, Hong BS, Hui R, Johansson I, Liu Y, Nilsson M, Nedyalkova L, Nordlund P, Nyman T, Min J, Ouyang H, Park HW, Qi C, Rabeh W, Shen L, Shen Y, Sukumard D, Tempel W, Tong Y, Tresagues L, Vedadi M, Walker JR, Weigelt J, Welin M, Wu H, Xiao T, Zeng H, Zhu H. In situ proteolysis for protein crystallization and structure determination. *Nat Methods* 2007;4(12):1019–1021. [PubMed: 17982461]
5. Longenecker KL, Garrard SM, Sheffield PJ, Derewenda ZS. Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of Rho-GDI. *Acta Crystallogr D Biol Crystallogr* 2001;57(Pt 5):679–688. [PubMed: 11320308]
6. Lawson DM, Artymiuk PJ, Yewdall SJ, Smith JM, Livingstone JC, Treffry A, Luzzago A, Levi S, Arosio P, Cesareni G, et al. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* 1991;349(6309):541–544. [PubMed: 1992356]
7. Kim Y, Quartey P, Li H, Volkart L, Hatzos C, Chang C, Nocek B, Cuff M, Osipiuk J, Tan K, Fan Y, Bigelow L, Maltseva N, Wu R, Borovilos M, Duggan E, Zhou M, Binkowski TA, Zhang RG, Joachimiak A. Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nat Methods* 2008;5(10):853–854. [PubMed: 18825126]
8. Vekilov PG. Solvent entropy effects in the formation of protein solid phases. *Methods Enzymol* 2003;368:84–105. [PubMed: 14674270]
9. Vekilov PG, Feeling-Taylor AR, Yau ST, Petsev D solvent entropy contribution to the free energy of protein crystallization. *Acta Crystallogr D Biol Crystallogr* 2002;58(Pt 10 Pt 1):1611–1616. [PubMed: 12351872]
10. Derewenda ZS. Rational protein crystallization by mutational surface engineering. *Structure* 2004;12(4):529–535. [PubMed: 15062076]
11. Cooper DR, Boczek T, Grelewski K, Pinkowska M, Sikorska M, Zawadzki M, Derewenda Z. Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallogr D Biol Crystallogr* 2007;63(Pt 5):636–645. [PubMed: 17452789]
12. Derewenda ZS, Vekilov PG. Entropy and surface engineering in protein crystallization. *Acta Crystallogr D Biol Crystallogr* 2006;62(Pt 1):116–124. [PubMed: 16369101]
13. Morgunkova A, Barlev NA. Lysine methylation goes global. *Cell Cycle* 2006;5(12):1308–1312. [PubMed: 16760670]
14. Nimura K, Ura K, Shiratori H, Ikawa M, Okabe M, Schwartz RJ, Kaneda Y. A histone H3 lysine 36 trimethyltransferase links Nkx2–5 to Wolf-Hirschhorn syndrome. *Nature* 2009;460(7252):287–291. [PubMed: 19483677]
15. Schmidt MW, Baldrige KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, John A, Montgomery J. General atomic and molecular electronic structure system. *J Comput Chem* 1993;14(11):1347–1363.
16. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983;4(2):187–217.
17. Pranata J, Wierschke SG, Jorgensen WL. OPLS potential functions for nucleotide bases. Relative association constants of hydrogen-bonded base pairs in chloroform. *J Am Chem Soc* 1991;113(8):2810–2819.
18. Feller SE, MacKerell AD. An improved empirical potential energy function for molecular simulations of phospholipids. *J Phys Chem B* 2000;104(31):7510–7515.
19. Woolf TB, Roux B. Conformational flexibility of o-phosphorylcholine and o-phosphorylethanolamine: a molecular dynamics study of solvation effects. *J Am Chem Soc* 1994;116(13):5916–5926.

20. Zhang M, Vogel HJ. Determination of the side chain pKa values of the lysine residues in calmodulin. *J Biol Chem* 1993;268(30):22420–22428. [PubMed: 8226750]
21. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79(2):926–935.
22. Darden T, York D, Pedersen L. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 1993;98(12):10089–10092.
23. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 1977;23(3):327–341.
24. James CP, Rosemary B, Wei W, James G, Emad T, Elizabeth V, Christophe C, Robert DS, Laxmikant K, Klaus S. Scalable molecular dynamics with NAMD 2005;26:1781–1802.
25. Pickett SD, Sternberg MJ. Empirical scale of side-chain conformational entropy in protein folding. *J Mol Biol* 1993;231(3):825–839. [PubMed: 8515453]
26. Sternberg MJ, Chickos JS. Protein side-chain conformational entropy derived from fusion data—comparison with other empirical scales. *Protein Eng* 1994;7(2):149–155. [PubMed: 8170918]
27. De Simone A, Dodson GG, Verma CS, Zagari A, Fraternali F. Prion and water: tight and dynamical hydration sites have a key role in structural stability. *Proc Natl Acad Sci U S A* 2005;102(21):7535–7540. [PubMed: 15894615]
28. Shaw N, Cheng C, Tempel W, Chang J, Ng J, Wang XY, Perrett S, Rose J, Rao Z, Wang BC, Liu ZJ. (NZ)CH...O contacts assist crystallization of a ParB-like nuclease. *BMC Struct Biol* 2007;7:46. [PubMed: 17617922]
29. Taylor R, Kennard O. Crystallographic evidence for the existence of CH...O, CH...N and CH...Cl hydrogen bonds. *J Am Chem Soc* 1982;104(19):5063–5070.
30. Vargas R, Garza J, Dixon DA, Hay BP. How strong is the C α -H...O=C hydrogen bond? *J Am Chem Soc* 2000;122(19):4750–4755.
31. Scheiner S, Kar T, Gu Y. Strength of the C α -H...O hydrogen bond of amino acid residues. *J Biol Chem* 2001;276(13):9832–9837. [PubMed: 11152477]
32. Makarov V, Pettitt BM, Feig M. Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment. *Acc Chem Res* 2002;35(6):376–384. [PubMed: 12069622]
33. Raschke TM. Water structure and interactions with protein surfaces. *Curr Opin Struct Biol* 2006;16(2):152–159. [PubMed: 16546375]
34. Kundrot CE, Richards FM. Effect of hydrostatic pressure on the solvent in crystals of hen egg-white lysozyme. *J Mol Biol* 1988;200(2):401–410. [PubMed: 3373535]
35. Han Z, Guo L, Wang H, Shen Y, Deng XW, Chai J. Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. *Mol Cell* 2006;22(1):137–144. [PubMed: 16600877]
36. De Boer JAA, Reinhoudt DN, Harkema S, Van Hummel GJ, De Jong F. Thermodynamic constants of complexes of crown ethers and uncharged molecules and x-ray structure of the 18-crown-6. (MeNO₂)₂ complex. *J Am Chem Soc* 2002;124(15):4073–4076.
37. Hao M-H. Theoretical calculation of hydrogen-bonding strength for drug molecules. *J Chem Theory Comput* 2006;2(3):863–872.
38. Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D. Toward rational protein crystallization: a web server for the design of crystallizable protein variants. *Protein Sci* 2007;16(8):1569–1576. [PubMed: 17656576]
39. Eisenberg, D.; Kauzmann, W. The structure and properties of water. Oxford: Clarendon Press; 1969. p. xii+296, [3] p
40. Dunitz JD. The entropic cost of bound water in crystals and biomolecules. *Science* 1994;264(5159):670. [PubMed: 17737951]
41. Fersht, A. Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding. New York: W.H. Freeman; 1999. p. xxi+631 p
42. Grimm C, de Ayala Alonso AG, Rybin V, Steuerwald U, Ly-Hartig N, Fischle W, Muller J, Muller CW. Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein sex comb on midleg. *EMBO Rep* 2007;8(11):1031–1037. [PubMed: 17932512]

43. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J Mol Biol* 1998;280(1):1–9. [PubMed: 9653027]

Fig. 1. The bi-molecular complex of water with ethyl ammonium cation (a), N,N-dimethyl ethyl ammonium cation (b), and N,N-dimethyl ethyl ammonia (c). The model compounds represent the side chains of protonated lysine, protonated dmLys and neutral dmLys, respectively

Fig. 2. The torsional angle along a simple C–C single bond, where X and Y denote heavy atoms, falls into either one of the three states as shown from *left to right*: *gauche(+)*, *trans* and *gauche(-)*

Fig. 3. The torsional angles (χ) in Lys (*top*) and dmLys (*bottom*). Only torsional angles around single bonds between two non-hydrogen atoms are counted. Because of the threefold symmetry along the CE–NE single bond in Lys, the torsional angle on this bond is not taken into account; for dmLys, however, the threefold symmetry no longer exists, and this torsional angle must be calculated also. As a result, there are 4 torsional angles being calculated for Lys, and 5 for dmLys

Fig. 4. The local hydrogen bonding patterns near the methylation site, showing the methylation site, Lys138 and the two neighboring aspartate residues, Asp242 and Asp137. **a** depicts the vicinity Lys138 in the unmodified protein, taken from the MD simulation trajectory at around the 6 ns mark; **b** also depicts the unmodified protein, but taken at around 12 ns in the simulation; **c** depicts the methylated protein, taken around 7.5 ns in the simulation

Fig. 5. The distances between ϵ -nitrogen of Lys/dmLys138 and δ -oxygen atoms of Asp242 and Asp137 versus simulation time. **a** depicts the distance between NE of Lys/dmLys138 and OD1 of Asp242 versus simulation time; **b** depicts the distance between NE of Lys/dmLys138 and OD2 of Asp242; **c** depicts the distance between NE of Lys/dmLys138 and OD1 of Asp137, and **d** depicts the distance between NE of Lys/dmLys138 and OD2 of Asp137. The distances in the unmethylated protein are shown in black and the methylated protein in red. The unit of time axis is picosecond, and the unit of the distance axis is Å

Fig. 6. The water density, radial distribution and entropy distribution near the Lys and dmLys side chain. **a** The bar chart showing the number of water atoms in consecutive 1 Å shells centered at ϵ -nitrogen of the lysine residue and the solid lines depict the corresponding water radial distribution function $g(O-O)$. Data from the unmethylated lysine are shown in black, the methylated lysine in red. The gray dashed line is the calculated water density assuming a uniform distribution, which is proportional to the square of the shell radius. **b** The water entropy distribution plotted with the density distribution (black unmethylated lysine; red methylated lysine). In the dmLys system, at between 3 and 6 Å from the residue side chain, there is a large

zone filled with low entropy water molecules, which is not present in the Lys system. Also in the dmLys system, this 'low entropy zone' has higher-than-normal water density

Fig. 7. Cartoon diagram of different crystal packing of the same protein before and after methylation on Lys138, which are shown in spheres. **a** The crystal packing of the unmethylated protein 2HO5, with Lys138 being far apart and the loop region between residues 154 and 171 tend to disrupt the crystal contact surface. **b** The methylated protein, 2HO3, packs in a different manner, with two dmLys forming head-to-head contact. The same loop region is now nestled in the V-shaped valley formed by the two helices, thus forming a more stable packing

Table 1

The side chain conformational entropy of methylated and unmethylated lysine residues

	S_{native}	S_{meth}	ΔS
Lys138A	8.31	11.81	3.50
Lys138B	9.48	12.97	3.45
Free	23.20	27.50	4.30

All units in $\text{J mol}^{-1} \text{K}^{-1}$