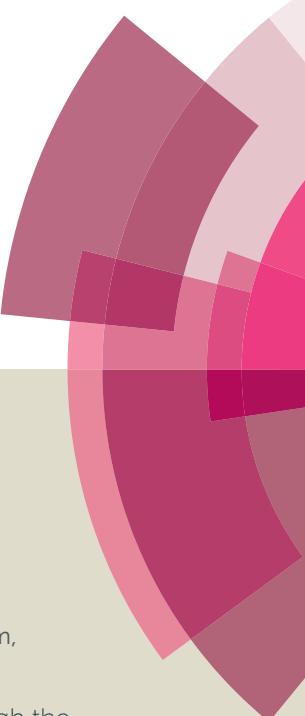
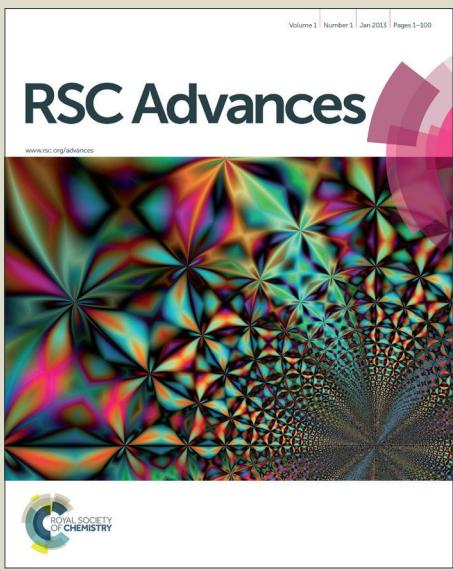


RSC Advances



This article can be cited before page numbers have been issued, to do this please use: S. Panneerselvam, P. Durai, D. Yesudhas, A. Achek, H. Kwon and S. Choi, *RSC Adv.*, 2016, DOI: 10.1039/C6RA16343B.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. This Accepted Manuscript will be replaced by the edited, formatted and paginated article as soon as this is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

**Journal Name****ARTICLE****Cysteine redox state plays a key role in the inter-domain movements of HMGB1: A molecular dynamics simulation study**

Suresh Panneerselvam, Prasannavenkatesh Durai, Dhanusha Yesudhas, Asma Achek, Hyuk-Kwon Kwon & Sangdun Choi*

Department of Molecular Science and Technology, Ajou University, Suwon 443-749, Korea

***Correspondence**

Sangdun Choi

Department of Molecular Science and Technology, Ajou University, Suwon 443-749, Korea

Fax: +82-31-219-1615

Tel: +82-31-219-2600

E-mail: sangdunchoi@ajou.ac.kr

ARTICLE

Abstract

High mobility group box protein 1 (HMGB1) is an abundant, conserved, non-histone nuclear protein that can serve as an alarmin, driving the pathogenesis of inflammatory and autoimmune diseases. In addition to its intracellular functions, HMGB1 can be released to the extracellular environment where it mediates the activation of the innate immune response, resulting in chemotaxis and cytokine release. HMGB1 contains three conserved redox-sensitive cysteines (C23, C45, and C106), and modifications of these cysteines determine the bioactivity of extracellular HMGB1. To advance our understanding of the redox-dependent functional changes of HMGB1, we have modeled full-length HMGB1 and simulated three different states of the protein, including its C23A and C106A mutants. Principal component analysis suggests that redox states affect the disordered loop movements, and subsequently the domain movements, of the active B-box domain that determines the fate of cytokine activity. We have also explored the free energy landscape of the redox states of HMGB1 to understand their crucial structural differences. These findings may have identified redox-dependent features that enable functional conformational transitions. Furthermore, active HMGB1 was docked onto a complex of Toll-like receptor 4 and myeloid differentiation factor 2 to predict the interactions that may provide helpful insights into the potential role of HMGB1 as therapeutic target for numerous autoimmune diseases.

Keywords: HMGB1, TLR4/MD2, damage-associated molecular patterns, Principal component analysis, DCCM

Introduction

High mobility group (HMG) proteins is a term given to a set of non-histone nuclear proteins with high electrophoretic mobility¹⁻³. The HMG proteins include three superfamilies termed HMGB, HMGN, and HMGA. High mobility group box protein 1 (HMGB1) is an abundant, conserved, non-histone nuclear protein that has important biological activities, both inside and outside the cell^{4, 5}. Inside the cell, HMGB1 binds to DNA and regulates gene transcription, along with several other functions. For instance, *in vivo* studies have shown that knocking out the HMGB1 gene results in the death of new born mice affected with hypoglycemia, highlighting the crucial role that this protein plays in the regulation of gene transcription⁶. Upon cellular activation, injury or death, HMGB1 can be translocated out of the cell. In the extracellular environment, HMGB1 can serve as a damage-associated molecular pattern (DAMP), where it can stimulate the innate immune system either by itself, or as part of immunostimulatory complexes with cytokines or other exogenous/endogenous molecules⁷. The biological activities of HMGB1 depends on the location, context and post-translational modification state of the protein². HMGB1 has a broad repertoire of immunological activities, playing multiple roles in the pathogenesis of inflammatory and autoimmune diseases, and mediating processes such as repair¹. These activities reflect the function of HMGB1 as an alarmin, and its ability to engage diverse receptors, including Toll-like receptors (TLRs) such as TLR2, TLR4, and TLR9⁸. During these interactions, post-translational modifications of HMGB1 can influence the receptor binding and downstream signaling events^{4, 9}. These post-translational modifications include acetylation, phosphorylation, methylation, and changes in the redox state of cysteine residues.

The structure of HMGB1 contains 215 amino acids, with two DNA binding domains (A- and B-boxes) and a C-terminal tail that contains a string of glutamic and aspartic acids³. Investigation of truncation mutants of HMGB1 identified that the B-box domain preserves the cytokine activity of HMGB1, whereas recombinant A-box domain antagonizes the function of the B-box domain¹⁰. HMGB1

ARTICLE

Journal Name

has two nuclear localization sequences (NLSs), one located in the A-box (residues 27–43) and the other in the B-box (residues 179–185). Four and five conserved lysine residues are present in NLS1 and NLS2, respectively¹¹. The lysine residues in the NLSs are susceptible to acetylation, resulting in the nuclear exclusion and consequent release of HMGB1. HMGB1 has three conserved redox-sensitive cysteines residues (C23, C45, and C106). Modification of these cysteine residues influences the bioactivity of the extracellular form of HMGB1. The cytokine-stimulating activity of HMGB1 depends on C23 and C45 being in a disulfide linkage. At the same time, C106 must be maintained in reduced form, as a thiol. This distinctive molecular conformation enables HMGB1 to bind to the TLR4/myeloid differentiation factor 2 (MD2)-complex, thereby signaling to induce cytokine release¹². When the cysteine residues in HMGB1 are in fully reduced form, HMGB1 cannot activate the TLR4/MD2 signaling pathway⁹. Likewise, complete oxidation of those cysteine residues causes HMGB1 to lose its immune-modulating activity. A similar effect is observed following the substitution of C23, C45, or C106 with either alanine or serine⁹. These studies reveal that post-translational modifications of HMGB1 determine its role in inflammation and immunity.

Even though the understanding of the function of HMGB1 has advanced steadily during the last decade, many of the extracellular biological functions of HMGB1 remain poorly understood³. At present, the stability of different forms of HMGB1, as well as their fates, remains unknown. The redox-dependent functional switching redox of HMGB1, in cells and in the extracellular space, has sparked interest in the discovery of new biological roles for this protein. To advance our understanding of the redox-dependent functional changes of HMGB1, we carried out molecular dynamics (MD) simulations for different states of HMGB1, namely; HMGB1 containing a disulfide linkage ($\text{HMGB1}^{\text{S-S}}$), HMGB1 with fully reduced cysteines ($\text{HMGB1}^{\text{SH-SH}}$), and HMGB1 with sulfonyl cysteines ($\text{HMGB1}^{\text{SO3H}}$). In addition, MD simulations were carried out for the HMGB1 mutants; C23A ($\text{HMGB1}^{\text{C23A}}$) and C106A ($\text{HMGB1}^{\text{C106A}}$). Furthermore, $\text{HMGB1}^{\text{S-S}}$ was docked onto TLR4/MD2, since we reasoned that the predicted interactions

might provide helpful insights into the potential of HMGB1 as therapeutic target for numerous autoimmune diseases.

Materials and Methods

1. Modeling of HMGB1. The full length of HMGB1 was built using I-TASSER^{13, 14}. I-TASSER server is an online web server for protein structure prediction. It allows academic users to automatically generate 3D structures of macromolecules. C-score is a confidence score to estimate the quality of the models developed by I-TASSER. C-score is typically in the range of [-5, 2], where higher C-score values signify quality models. Disulphide bond between C23 and C45 of HMGB1^{S-S} was built using Modeller software¹⁵. Modeller is widely used for comparative modeling of protein three-dimensional structures. The user provides an alignment of a target sequence and the known related structures as templates to model a protein. The HMGB1^{SO3H} was built using Molecular Operating Environment (MOE; Chemical Computing Group Inc., Montreal, Canada). All the models can be downloaded from the figshare link¹⁶.

2. Molecular Dynamic simulations

Simulations were performed for HMGB1^{S-S}, HMGB1^{SH-SH}, HMGB1^{SO3H}, HMGB1^{C23A}, HMGB1^{C45A} and HMGB1^{C106A} using GROMACS 5.0 software¹⁷. The parameters were assigned for proteins using an AMBER99SB-ILDN force field¹⁸. A cubic box was set up by specifying a distance of 10 Å between the protein and the box edge. The systems were then solvated using the TIP3P water model¹⁹, and were neutralized by adding an appropriate number of counter-ions. All bonds were constrained using the LINCS algorithm²⁰, allowing an integration time step of 2 fs. The Verlet cutoff scheme²¹ was used with a minimum cutoff of 10 Å for short-range Lennard-Jones interactions, and with real-space contributions to

ARTICLE

the smooth particle mesh Ewald algorithm²², which was used to compute long-range electrostatic interactions. A stepwise protocol was employed for equilibration, beginning with a 500 ps simulation under constant volume (NVT) conditions, followed by a further 500 ps switching to constant pressure (NPT) conditions. All production simulations were performed with a 2 fs time step, and the coordinates were saved every 2 ps under constant pressure (1 bar) and temperature (300 K) without any position restraints. Production simulations were carried out for 100 ns at constant temperature (300 K) on the redox state models of HMGB1^{S-S}, HMGB1^{SH-SH}, HMGB1^{SO3H}, HMGB1^{C23A}, and HMGB1^{C106A}.

3. Protein-protein interactions

Docking was performed by taking the representative lowest energy conformation of the HMGB1^{S-S} model (100 ns simulation) from the free energy landscape (FEL) analysis, and by taking the X-ray crystal structure of the human TLR4/MD2 complex (PDB code: 3FXI). Hetero atoms were removed from the TLR4/MD2 complex. As an initial step, the protein structure was prepared by assigning hydrogen atoms, followed by brief energy minimization. The residues at the TLR4/MD2 protein interface were predicted using InterProSurf²³, which predicts residues in proteins that are most likely to interact with other proteins. InterProSurf can be used most efficiently to locate functionally important sites on the protein surface, when combined with evolutionary information on protein sequences, and with data from mutagenesis experiments. The predicted protein-protein interactions interface residues for TLR4 were: M41, E42, F63, D84, R87, T110, V132, E135, H159, S183, L212, K230, R234, R243, F263, R264, N265, R289, V316, V338, and N339. The predicted interface residues for MD2 were: Y42, I66, R68, M85, N86, L87, P88, R90, R96, S98, D99, D100, D101, Y102, S103, R106, L108, K109, G110, E111, T112, T115, T116, S118, G123, I124, K125, and S127. Sites known to interact from experiments were also used to guide the docking process. The active sites residues in the TLR4/MD2 complex were defined as follows: R264, E439, K341, K362, S416, N417, F440, L444, and F463 for TLR4. K58, V82, M85, L87, R90, S118, K122, G123, I124, K125, and F126 for MD2²⁴. The TLR4 binding domain was used as an active

site for HMGB1^{S-S}. HADDOCK^{25, 26} is a docking method that works based on the available experimental knowledge about interface regions among molecular components and their relative orientations. Compared to other docking programs, HADDOCK allows the conformational changes in molecules during complex formation. HADDOCK has also performed well in several blind docking experiments and is currently the most cited biomolecular docking program. The neighbouring residues of all structures were considered as passive residues for this docking protocol. The best docked cluster was selected based on the HADDOCK score and the RMSD. The MD simulation procedure described above was also applied to TLR4/MD2/HMGB1^{S-S} using a 25 ns production run.

4. Principal component analysis and Free energy analyses

Principal component analysis (PCA) is a statistical method which can be used to describe the most relevant correlated motions, using a new basis set directly reflecting the collective motions undergone by the system. A detailed description of the method is found elsewhere^{27, 28}. High-amplitude, concerted motion in the protein trajectories, through the eigenvectors of the covariance matrix of protein atomic fluctuations, can be unveiled using PCA. The first 20 projection eigenvectors of the protein were extracted from simulated complexes and analyzed for their cosine content. The first two eigenvectors (PC1 and PC2) having cosine content less than 0.2 were used to define the FEL analyses. The analyses were done using the bio3d²⁹ module in the R analytic software tool. The g_sham module of the GROMACS package was used to calculate the FEL. Contour maps of the FEL were generated using a trial version of Mathematica (Wolfram Research, USA).

5. Dynamical cross-correlation matrix (DCCM) calculations

The time correlated motions (DCCM) of backbone atoms during simulation of the proteins were calculated over the entire trajectory, using the bio3d module of the R analysis tool. A snapshot was taken at every 50 ps interval of the trajectory, and a covariance matrix was generated between residues *i* and *j*.

ARTICLE

Before the generation of the covariance matrix, overall translational and rotational motions of the protein were removed. A cross-correlation coefficient was calculated for backbone atoms.

6. Analyses

The method we followed to determine the binding free energy, following the MM/PBSA approach, has been described previously^{30, 31}. In this work, Δ polar and Δ nonpolar values were calculated with APBS³². The GMXAPBS analysis tool³³ in GROMACS was used to predict the binding free energy. Structures were selected every 100 ps from the last stable 10 ns simulation trajectory of the 25 ns simulation of the docked TLR4/MD2/HMGB1^{S-S}. Instead of performing simulations on the single mutant complexes, it is possible to perform binding free energy calculations of alanine mutant complexes using the MM/PBSA approach on snapshots taken from the wild-type simulation. We performed the *in silico* alanine scanning on HMGB1^{S-S} using the TLR4 binding domain in the docked complex. Tools in GROMACS were used for the trajectory analyses. All visualizations were performed using Chimera³⁴ (UCSF, San Francisco, USA) and PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC).

Results

1. Modeling the isoforms of HMGB1

The protein sequence of human HMGB1 was retrieved from UniProt (ID: P09429) and queried against the Protein Data Bank (PDB) using the PSI-BLAST server. The results showed a 77% identity with the human tandem HMG box domain, whose solution-structure had been determined by a structural genomics/proteomics initiative (PDB code: 2YRQ). Most of the sequence in the C-terminal tail presented no similarity. Therefore, we used the I-TASSER server to model the full-length protein¹³. The I-TASSER server is one of the most widely used online systems for automated protein structure prediction and structure-based functional annotation¹⁴. The core programs have been extensively tested in benchmarked,

blinded experiments, which have established the advantage of the I-TASSER server over other state-of-the-art methods. The I-TASSER server used the template structure (PDB code: 2YRQ) for model building, and for consensus agreement using the PSI-BLAST server. Out of 215 amino acids, 173 matched the structure of HMGB1. For the remaining C-terminal tail regions (residues 174–215), secondary structure prediction suggested them to be coiled regions. The I-TASSER model was chosen based on “C-score” (Fig. 1a). A C-score is a reliability score used to estimate the quality of models predicted by I-TASSER. It should be noted here that a few structures of A- and B-box domains are available. These structures have been solved separately, in complex with DNA, or in complex with other proteins^{35–38}. The HMGB1 structure consists of an A-box DNA binding domain (residues 1–79), linker loop 1 (residues 80–88), a B-box DNA binding domain (residues 89–162), linker loop 2 (residues 163–185), and the C-terminal tail (residues 186–215). There are two NLS signal domains, NLS1 (residues 27–43) and NLS2 (residues 178–186). The important TLR4 binding domain is located in the B-box domain (residues 89–108). One of the important residues, C106, is located in the TLR4 binding domain. On the other hand, the other two key residues, C23 and C45, are located in the A-box DNA binding domain. HMGB1 is a helical protein with three helices in each of the A- and B-box domains. The remaining C-terminal tail is a coiled region including a small helical region. Hereafter, we will take residues 1–79 to be the A-box DNA binding domain, residues 80–88 to be linker loop 1, residues 89–108 to be the TLR4 binding domain, residues 89–162 to be the B-box DNA binding domain, residues 163–185 to be linker loop 2, and residues 186–215 to be the C-terminal tail. Intrinsically disordered regions have been reported in HMGB proteins³⁹; therefore we sought to predict them. Those regions have a functional role, such as the basic disordered C-terminal tail, which becomes structured upon binding to DNA. The consensus disordered regions predicted using the MetaDisorder⁴⁰ server suggested that the N-terminal region of the protein (residues 1–12), and parts of the A-box domain, B-box domain, and the C-terminal tail contain disordered regions, including part of TLR4 binding domain (Fig. 1b). Next, we used the full-length structural model of HMGB1, building a disulfide bond between C23 and C45 using Modeller⁴¹. The

ARTICLE

sulfonyl cysteines and other mutant proteins were prepared using MOE. The full-length of structures of HMGB1 in different redox states were energy minimized fully, and simulated for 100 ns, for further study. The validation of modeled active and inactive HMGB1 states was reasonably good and the scores are presented in Table S1.

2. Fully reduced HMGB1 becomes destabilized, leading to increased flexibility

To gain insight into the mechanism of HMGB1 activity, we simulated (100 ns) three different redox states of HMGB1; namely $\text{HMGB1}^{\text{S-S}}$, $\text{HMGB1}^{\text{SH-SH}}$, and $\text{HMGB1}^{\text{SO3H}}$. The $\text{HMGB1}^{\text{C23A}}$, $\text{HMGB1}^{\text{C45A}}$ and $\text{HMGB1}^{\text{C106A}}$ mutants were also simulated. The isoforms of HMGB1, MD2 binding, and cytokine/chemokine activity are shown in Fig. 1c. Fig. 2a shows the root mean square deviations (RMSD) of the structures sampled during MD simulations, from their respective starting structures. Examination of the data presented in the RMSD plots shows that $\text{HMGB1}^{\text{S-S}}$ was stable from about 40 ns during the production phase of the MD simulations. This result was anticipated, owing to the disulfide bond between C23 and C45. The trajectory for $\text{HMGB1}^{\text{SH-SH}}$ showed that the equilibrated structure deviated by $\sim 10 \text{ \AA}$ from the starting structure. When compared to $\text{HMGB1}^{\text{S-S}}$, the structure had deviated by $\sim 5 \text{ \AA}$. The major structural changes occurred during the initial 20 ns, with a sudden increase in the RMSD to $\sim 10 \text{ \AA}$. The structural deviation then proceeded rather slowly over the rest of the simulation, reaching a final value of $\sim 15 \text{ \AA}$. In the case of $\text{HMGB1}^{\text{SO3H}}$, the major structural changes occurred during the initial 1–5 ns, remaining relatively stable thereafter. In the case of $\text{HMGB1}^{\text{C106A}}$, the structure was dynamic throughout the simulation, similar to $\text{HMGB1}^{\text{SH-SH}}$. In contrast, in the case of $\text{HMGB1}^{\text{C23A}}$, the trajectory did not stabilize until 40 ns, remaining stable thereafter. It is apparent from the results that the $\text{HMGB1}^{\text{SH-SH}}$ and $\text{HMGB1}^{\text{C106A}}$ structures were dynamic throughout the simulation. The trajectories of $\text{HMGB1}^{\text{SO3H}}$ and $\text{HMGB1}^{\text{C23A}}$ were rather similar, with comparable RMSD profiles. The absence of the C23-C45 disulfide bond led to an increased flexibility in the $\text{HMGB1}^{\text{SH-SH}}$, $\text{HMGB1}^{\text{SO3H}}$, $\text{HMGB1}^{\text{C23A}}$, $\text{HMGB1}^{\text{C45A}}$, and $\text{HMGB1}^{\text{C106A}}$ simulated models.

Root mean square fluctuation (RMSF) analyses are depicted in Fig. 2b. From these data, it is clear that for all the proteins the RMSF is indeed higher than that of HMGB1^{S-S}, except for the HMGB1^{C23A} model. HMGB1^{C23A} had similar fluctuations to HMGB1^{S-S}. Notable differences were observed in the fluctuations of HMGB1^{C23A} in two locations; near the TLR4 binding domain, and in the C-terminal tail. For the HMGB1^{SH-SH}, HMGB1^{SO3H}, and HMGB1^{C106A} models, higher fluctuations were observed across all residues when compared to the HMGB1^{S-S} model. From these data, it is clear that the RMSF of the inactive HMGB1 species are indeed higher than that of active HMGB1. Specifically, higher flexibility was observed in the TLR4 binding domain, as well as the B-box domain of HMGB1^{SH-SH}. It is well known that the TLR4 binding domain is crucial for cytokine inducing activity, and that this region undergoes significant conformational changes. We also computed the radius of gyration (R_g) as a function of time, because any destabilization of the protein structure would result in a large increase in R_g values. The R_g was calculated for the models, as shown in Fig. 2c. For HMGB1^{S-S}, a relatively steady R_g was maintained throughout the simulation, suggesting that form of the protein is stable. In contrast, drastic R_g changes were observed for the other proteins modeled. In particular, the R_g plots for HMGB1^{SH-SH} showed an increase of ~6 Å.

To address the potential role played by the disulfide bond in the activity of HMGB1, we assessed whether its presence influenced the secondary structure of HMGB1 through simulations of the protein bearing single point mutations. Unusually, the secondary structure was comparable with that of HMGB1^{S-S}, except for a few regions. The C23-C45 bond is positioned between the two helices (I and II) in HMGB1. We also monitored the distance between the cysteines. The closest distance attained during the simulation is shown in Fig. 2d. As expected, the disulfide bond in HMGB1^{S-S} was stable throughout the simulation. At the beginning of the simulation on HMGB1^{SH-SH}, the distance between C23 and C45 was about 4 Å, and after fluctuating somewhat, it finally stabilized. These results clearly indicate that the disulfide bond structurally stabilizes the protein. However, in the case of HMGB1^{SO3H}, the distance

ARTICLE

Journal Name

between C23 and C45 was around 3 Å, and that distance remained stable throughout the simulation. As reported previously, any mutation in one of the cysteine residues significantly reduces the activity of HMGB1; thus it was suggested that this disulfide bond makes a structural contribution to the ability of the protein to function normally during its activation mechanism¹⁰.

3. A disulfide bond maintains the inter-domain movements during the activity of HMGB1, and may protect HMGB1 from destabilization events

To explore dynamic movements we evaluated dynamical cross-correlation maps (DCCMs) of backbone atoms for all the complexes (Fig. S1a–f). There was a clear domain decomposition observed for HMGB1^{S-S}, with both correlated and anti-correlated motion, in contrast to HMGB1^{SH-SH}, HMGB1^{SO3H}, HMGB1^{C23A}, HMGB1^{C45A}, and HMGB1^{C106A} complexes. In the case of HMGB1^{SH-SH}, part of the A- and B-box domains, along with linker loop 2, exhibited anti-correlated motions, in contrast to HMGB1^{S-S}. The main difference was observed in an increased level of anti-correlated motion for HMGB1^{SH-SH} and HMGB1^{SO3H}, when contrasted with HMGB1^{S-S}. Specifically, the TLR4 binding domain HMGB1^{S-S} showed significant differences from the other models, suggesting differences in domain motion.

In addition, PCA was carried out using the MD simulation trajectories. PCA highlighted the differences in motion of the following complexes: the active HMGB1^{S-S}, and the inactive HMGB1 species comprising HMGB1^{SH-SH}, HMGB1^{SO3H}, HMGB1^{C23A}, HMGB1^{C45A}, and HMGB1^{C106A}. PCA identifies appropriate low energy displacements of groups of residues, and highlights the amplitude and direction of dominant protein motions by projecting the trajectories onto a reduced dimensionality space, thus distilling the slow modes captured in the trajectories²⁷. These collective motions represent the critical biological motions that determine the functional state of a protein. To gain insight into functional significance, we generated movies for all the complexes to visualize the motions of the three dominant PCs (Movies S1 to S6). The first three PCs cover 70.9% of the overall motion of HMGB1^{S-S} (Fig. S2a). Both components are dominated by internal motions, as well as the overall translational motion of the

protein. In the case of HMGB1^{S-S}, the A- and the B-box domains move in opposite directions (i.e. the A-box domain rotates clockwise and B-box domain rotates anti-clockwise). Both of the PCs involve substantial motions of linker loop 1 and the TLR4 binding domain, which leads to the movement of the A- and B-box domains in opposite directions (Movie S1). We note that the motion of the TLR4 loop is necessary, and this may be the reason for the activity of HMGB1^{S-S}. The global shape and conformation of HMGB1^{S-S} remain intact, which likely modulates binding interactions with its receptors. This analysis suggests that active HMGB1^{S-S} undergoes coupled, non-contact domain movement⁴².

The first three PCs cover 78.9% of the overall motion of HMGB1^{SH-SH} (Fig. S2b). Owing to absence of a disulfide bond in HMGB1^{SH-SH}, helix I and helix II were not intact in the A-box domain, which leads to prominent and dramatic movement of linker 1 loop, and affects the organized B-box domain in turn (Movie S2). In particular, the orientation of the B-box domain, together with the lack of movement in the TLR4 binding domain loop, were also different when compared to HMGB1^{S-S}. Flexibility was observed in both the N-terminus and the C-terminal tail in HMGB1^{SH-SH}. The first three PCs cover 81% of the overall motion in HMGB1^{SO3H} (Fig. S3a). When the B-box domain in HMGB1^{SO3H} was compared to that in HMGB1^{SH-SH}, it was found to adopt an opposite orientation, and to have different motions. The A-box domain of HMGB1^{SO3H} remained intact. The TLR4 domain in HMGB1^{SO3H} adopts yet a different orientation, and it moves differently, leading to an inward movement of the B-box domain. In addition, movement was observed in the N-terminal loop (Movie S3). The first three PCs cover 63% of the overall motion in HMGB1^{C23A} (Fig. S3b). HMGB1^{C23A} had N-terminal movement, and both the orientation and motions of its B-box domain were affected (Movie S4). Alterations were also observed in the conformations of the B-box domain. In case of C45A, the first three PCs cover 59.06%, which is different from rest of the HMGB1 simulated states (Fig. S4a). In the case of C106A, first three PCs cover 85% of the overall motion in HMGB1^{C106A} (Fig. S4b). In HMGB1^{C106A}, the dramatic movement of the A-box domain induced the outward movement of the B domain (Movie S6). It is clear from the scatter plot

ARTICLE

(Fig. S2–S4) that the eigenvectors computed from the MD trajectories of the systems are quite varied, which indicates clearly the differences in protein motion between the active and inactive redox forms of HMGB1. These results further confirm that there are differences in the inter-domain movements between active and inactive forms of HMGB1. Thus, the differences between the various structures sampled during the simulation arise from inter-domain movements.

4. Free energy landscape analyses of conformational changes in HMGB1

The energy landscape theory of protein folding advances the understanding that the mechanism of folding is regulated by the formation of native contacts, leading to a funnel-shaped energy landscape, in which energy decreases with increasing formation of the native structure⁴³. The FEL offers a valuable resource for understanding different conformation states in the folding process and their pathways of interconversion⁴⁴. The dynamics responsible for protein conformational changes, are governed by the properties of the conformational energy landscape. These conformational changes range from large-scale protein folding to smaller changes, such as those achieved by different redox states. To obtain the FEL, we performed PCA on the five simulated structures ($\text{HMGB1}^{\text{S-S}}$, $\text{HMGB1}^{\text{SH-SH}}$, $\text{HMGB1}^{\text{SO}_3\text{H}}$, $\text{HMGB1}^{\text{C}^{23}\text{A}}$, $\text{HMGB1}^{\text{C}^{45}\text{A}}$, and $\text{HMGB1}^{\text{C}^{106}\text{A}}$), and determined the two principal axes that span the conformational space. A FEL plot was then generated, using PC1 and PC2, to understand the conformational changes that the structure of $\text{HMGB1}^{\text{S-S}}$ undergoes during the course of simulation (Fig. 3). There were two major basins observed, as well as a minor basin ($\text{HMGB1}^{\text{S-S}}$). The regions of the conformational space corresponding to these basins are referred to as region I, II and III. The depth of an energy minimum represents the thermodynamic stability of a protein; the heights of barriers separating energy minima dictate the kinetic stability of a protein, that is, how readily it can leave one conformation and sample another; and the width of an energy minimum correlates with the breadth of the conformational ensemble within the energy well⁴⁵. The FEL suggests that the changes were relatively local in nature, occurring in turns and loops. Changes were observed in the orientations of the B-box

domain arising from differences in loop movements. Different sub-states have the same overall structure but they differ in detail. For instance, they perform the same function, perhaps with different rates⁴⁶. It seems likely from these results that the topology of the domains is stable and that the main changes are in the loop regions, including the functionally important TLR4 binding domain. Considering that loops are disordered regions, we can expect them to adopt different conformations. The multi-conformational sub-states instantiated by loop motions provide opportunities for the protein to interact with multiple, structurally dissimilar partners of functional importance⁴⁷. The structure found to be lowest in the major basin was used as a reference structure for further studies.

In the case of HMGB1^{SH-SH}, the regions of the conformational space corresponding to these basins are referred to as region I, II, III, and IV (Fig. 4). As can be seen, there are clearly observable changes in the conformation of the structure. One clear pathway leading to unfolded structures can be found by looking at the distribution of configurations, from state I to state IV. Hierarchical landscapes characterize the dynamical behavior of proteins, which in turn depends on the relationship between the topology of the basins, their transition paths, and their kinetics over energy barriers⁴⁸. It appears from the structures from the FEL plots, as well as from the movie (Movie S2), that without the constraint of the disulfide bond the direction of domain A and B-box domains disorganized, resulting in events that destabilize the protein. The variable loops of the protein structures allow both A and B domains to move in opposite directions, leading to differences in domain orientation. In addition, the collapse of tertiary structures and the occurrence of changes in the N- terminus, were observed during the simulation. These alternative conformations may enable functional interactions by exposing interactive surfaces, providing opportunities for new, favorable interactions (e.g. chemokine activity). We showed that the loss of disulfide bond brings about a large destabilization, giving rise to dramatic changes in the structure of HMGB1^{SH-SH}. Therefore, the disulfide bond in HMGB1 is undoubtedly critical for stabilizing the structure. These findings may represent the general features of conformational transitions within the inactive state of HMGB1^{SH-SH}.

ARTICLE

In the case of HMGB1^{S03H}, the transition that the B domain makes from the state I to IV, when its orientation turns inwards, is clearly visible (Fig. 5). The passage over four global states of HMGB1^{S03H} is clearly seen as the result of two principal modes of reconfiguration. The shape of HMGB1^{S03H} was altered dramatically by oxidation. Movement was also observed in the N-terminus, and there were many minor basins. Fig. 6 shows the FEL of HMGB1^{C23A}, which contains three energy basins. Two major changes were observed in the three energy basins. The change in the orientation of helix III in the A-box domain was clearly visible in all three structures. Further, changes in the orientation of the B-box domain, and the change in the TLR4 binding domain region in particular, were observed. When the structures of HMGB1^{S-S} and HMGB1^{C23A} were superimposed, it became clear that the A-box domain is affected by the mutation, and that this in turn affects the orientation of the B-box domain. The HMGB1^{C45A} has two major energy basins. The two structures are different and the loop regions affect the domain movement. The mutations HMGB1^{C23A} and HMGB1^{C45A} affect the structure quite differently (Fig. 7). In the case of HMGB1^{C106A}, three major energy basins were observed, as well as two minor energy basins (Fig. 8). Even a minor change in the TLR4 binding domain would cause a difference in the orientation and organization of the A and B domains. Mild mutation will initially shift the folding routes if one set of routes to the native structure is completely blocked by a very destabilizing mutation⁴⁹. When HMGB1^{C106A} was superimposed with active HMGB1^{S-S}, it became clear that the B-box domain was affected.

5. Model of the TLR4/MD2/HMGB1^{S-S} complex

Previous reports had demonstrated that extracellular TLR4/MD2 complexes bind specifically to the cytokine-inducing disulfide isoform of HMGB1^{S-S}⁵⁰. Therefore, we were interested in predicting the interaction between HMGB1^{S-S} and the TLR4/MD2 complex. Experimental studies have reported that in mutants of HMGB1, the A-box domain acts as an antagonist⁵¹ to HMGB1, whereas the B-box domain exerts its cytokine-inducing function¹⁰. Residues 89–108 contain the minimal sequence for the pro-

inflammatory activity of the B-box¹⁰. Hence, we have chosen those residues as the active site for HMGB1^{S-S}. Experimental studies have also reported that the active sites residues for TLR4/MD2 are defined as follows: In TLR4; R264, K341, K362, S416, N417, Q436, E439, F440, L444, and F463; and in MD2; K58, V82, M85, L87, R90, S118, K122, G123, I124, K125, and F126²⁴. It has been suggested that oligomerized HMGB1 binds to the TLR4/MD2 complex with a higher affinity than monomeric HMGB1^{52, 53}. However, our aim was to predict the interaction mode between TLR4/MD2 and monomeric HMGB1 by docking. Additionally, owing to the complexity of the simulation, we docked monomeric HMGB1 on to only one TLR4/MD2 heterodimer. It has been shown experimentally that the TLR4/MD2/HMGB1^{S-S} complex is capable of signaling, without needing a ligand such as lipopolysaccharide (LPS)³. Therefore we have used only the TLR4/MD2 complex for docking. TLR4 adopts the characteristic horseshoe shape of the leucine-rich repeat superfamily. MD2 adopts a structure with a β-cup fold composed of two anti-parallel β-sheets, which form a large hydrophobic pocket²⁴. Ten different clusters of docked TLR4/MD2/HMGB1^{S-S} complexes were suggested by HADDOCK²⁶. One important feature of the best ranked conformations in HADDOCK is that the interaction regions that are represented are similar to each other. The differences were only in the orientation of the A-box domain. The first three binding modes are shown in Fig. S5a–c.

Out of three conformations of active HMGB1^{S-S} models given in Fig. 3, two of them have similar binding conformations (Fig. S6). However, for TLR4/MD2/HMGB1^{S-S} model III, the orientation of binding is different. Since the loops are flexible in binding orientation, B box domain changes the binding pose. This may also be due to the change in loop conformation of HMGB1^{S-S}. The consensus and lowest energy structures among the docked complexes were selected and used for 25 ns MD simulations. In order to understand the binding modes of inactive HMGB1 redox states, we docked HMGB1^{SH-SH}, HMGB1^{SO3H}, HMGB1^{C23A}, HMGB1^{C45A}, and HMGB1^{C106A} and are shown in Fig. 9. The protein-protein interaction energies are presented in Table S2. Due to the redox states, the inter-domain movement of HMGB1 was affected and as a result, the binding conformations change drastically. The binding differences are clearly

ARTICLE

visible between active and inactive HMGB1 states. Furthermore, it seems from this analysis that the geometry of active site is essential for the active HMGB1^{S-S}. To obtain a FEL profile, we performed PCA and determined the two principal axes that span the conformational space (Fig. S7). The RMSD of the complex suggests that in last 10 ns of the simulation, the trajectory was comparatively stable. The lowest energy structures from three major basins of the FEL were superimposed, and these are shown in Fig. S7. The conformational change in the orientation A-box is clearly visible. After MD simulations, the lowest structure of the FEL was analyzed to identify residues within a radius of 3.5 Å from the complex that form hydrophobic or electrostatic interactions, or that form hydrogen bonds. Fig. 10a–d shows the interaction mode of TLR4/MD2 and HMGB1^{S-S}, and the interacting residues. The docking interaction suggests that there are hydrogen bonds, ionic interactions, and also a few hydrophobic interactions between TLR4/MD2 and HMGB1^{S-S} (Table 1). MD2 residues S28, S45, N47, Q53, K58, and N158 form hydrogen bonds with HMGB1^{S-S} residues K96, R110, S107, K96, Y96, Y155, and K173, respectively. Ionic interactions were observed between MD2 (Q53 and N158) and HMGB1^{S-S} (K96 and N173). Only one hydrophobic interaction was observed, between MD2 W23 and HMGB1^{S-S} F103. In case of TLR4 and HMGB1^{S-S}, the residues D238, Q266, N268, D294 and K146, K150, K147, K154, form hydrogen bonds. Moreover, the ionic interactions were observed between TLR4 residues (Q135, D238, Q266, D294), and HMGB1^{S-S} residues (K112, K146, K150, K147, K154), and there were no hydrophobic interactions within the docked complex. The hydrogen, ionic and hydrophobic interactions of TLR4/MD2 and HMGB1^{S-S} were observed during the MD simulation. The interactions were relatively stable (Fig. S8). In summary, docking simulations suggest that both hydrogen bonding and ionic interactions contributed to HMGB1 binding with TLR4/MD2.

We next investigated the energetic parameters driving the interaction between TLR4/MD2 and HMGB1^{S-S} by performing binding free energy calculations on the complex using a molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) approach³⁰. The MM/PBSA method calculates the free energy of binding as the difference between the free energy of the complex and the free energy of the

receptor and the ligand, averaged over a number of trajectory snapshots. To investigate the role of the TLR4 binding domain loop, we mutated the loop residues and recalculated the binding free energies. The results of these calculations are summarized in Table 2. The residues A94, P98, P99, and A101 were not used for alanine scanning. When the total binding energies were decomposed into individual components, we found the van der Waals, and non-polar/non-polar solvation interaction energy terms were the dominating factors holding the TLR4/MD2/HMGB1^{S-S} complex together. Dramatic changes in the free energy were observed in the complexes of HMGB1 mutants (C45A, D91A, N93A, R97A, and E108A) suggesting that these residues play a significant role in the formation of the complex. Another set of residues, such as C23, F89, K90, K96, S100, F102, F103, L104, and F105, make moderate contributions to assist complex formation. The only exception is the C106A mutant in the TLR4 binding domain, which has stronger binding affinity when mutated. Alanine scanning mutagenesis suggests that C106 is less likely to contribute to the TLR4/MD2/HMGB1^{S-S} complex. Taken together, these data build up a body of arguments in support of the roles played by non-polar and van der Waals interactions in maintaining the stability of the complex. Meanwhile, the binding was strongly antagonized by positive polar solvation contributions (Δ_{psol}) and Coulombic (Δ_{coul}) contributions. Overall, alanine mutations decreased polar solvation contributions with respect to the wild-type complex.

Discussion

HMGB1 plays a functional role as a signaling molecule that informs other cells that damage or invasion has occurred¹. Earlier studies have reported that the products of cellular injury activate fundamental defense mechanisms, which are equivalent to responses activated by molecules from pathogens³. HMGB1 is a molecule with many faces, showing different activities depending on its redox modifications, with or without hyperacetylation, during inflammation and apoptosis². An understanding of the conformational changes that occur in HMGB1 during different redox states are essential of an understanding of its

ARTICLE

function. Therefore, we generated models and performed MD simulations on HMGB1^{S-S}, HMGB1^{SH-SH}, and HMGB1^{SO3H}, as well as the mutants HMGB1^{C23A}, HMGB1^{C45A}, and HMGB1^{C106A}. The simulated trajectory suggests that, while the HMGB1^{S-S} structure was stable, the other models exhibited clear structural differences in both stability and flexibility. This observation is based on analyses of the RMSD, RMSF, and R_g values of the structures sampled during various simulations. A relatively steady R_g was maintained for HMGB1^{S-S}, suggesting that the disulfide bond stabilizes the A domain throughout the simulation, which also leads to a stable B domain. It is apparent that the disulfide bond plays a crucial role in the folding and structural stabilization of HMGB1^{S-S}. To further understand some of the interesting structural characteristics of the HMGB1 domain movements, we inspected structures that represent populated regions using PCA. Figs. S2-4 highlight the dominant changes in motion across two principal components in active and inactive HMGB1 redox states. It is observed that eigenvectors computed from individual trajectories were varied in all the systems, which further emphasize on differences in conformational landscapes between the active and inactive HMGB1 conformations. This difference in average eigen values across the first two principal components suggests a greater mobility of the inactive conformations as compared to the active HMGB1. This may be due to the disulphide bond between C23 and C45. The PCA also suggests that the main direction of the motion of the domain is different in the two proteins. Furthermore, our analysis suggested that the active HMGB1^{S-S} has coupled, non-contact domain movements in both the A- and B-box domains⁴². The disulfide bond effectively pins helix I and II in place, locking the movements of the A domain. Such positioning appears instrumental in maintaining the precise geometry of the active site in HMGB1^{S-S}, while allowing a functionally oriented modulation of the loops resulting from the relative motions of the two subdomains. It is plausible that the reason truncated mutants of the HMGB1 B-box protein preserve their cytokine activity might be due to the maintenance of the precise positioning of TLR4 binding domain loop in such mutants. In the case of HMGB1^{SH-SH}, HMGB1^{SO3H}, HMGB1^{C23A}, HMGB1^{C45A}, and HMGB1^{C106A}, intrinsic loop movements affect the domain movements, leading to a substantial loss of cytokine activity. This distinctive

conformation enables HMGB1^{S-S} to bind to the TLR4/MD2 complex, and thereby signal to induce cytokine release. The FEL of the conformational states of HMGB1^{S-S} suggest that the changes were mostly in the linker and loop regions. There was also a clear domain decomposition observed for HMGB1^{S-S}, with correlated and anti-correlated motions present, in contrast to the models of HMGB1^{SH-SH}, HMGB1^{SO3H}, HMGB1^{C23A}, HMGB1^{C45A}, and HMGB1^{C106A}. The superimposed structures of HMGB1^{S-S} and other isoforms of HMGB1 are shown in Fig. S9a–f. The presence of fully reduced cysteines clearly affected the B-box domain. Terminal oxidation of cysteines changes the rearrangement in the A- and B-box domains, which appear more compact. When the structures of stable HMGB1^{S-S} and the HMGB1^{C23A} mutant were superimposed, it was seen that the C23A mutation had affected the assembly of A-box domain. When the structures of HMGB1^{S-S} and the HMGB1^{C106A} mutant were superimposed, it was seen that the global shape had been affected by the C106A mutation in the TLR4 binding domain. This clearly suggests that shape of the molecule was affected, including the orientation of the A- and B-box domains. Other redox isoforms of HMGB1 do not bind to MD2, and therefore do not activate the TLR4 system. The reason for this is that the shape of the molecule is affected in these isoforms.

One interesting thing to point out is that the disordered regions in HMGB1 are the reason for its ability to engage diverse receptors, including TLR2, TLR4, TLR9, and RAGE³ (receptor for advanced glycation end products). Moreover, certain disordered regions might serve as molecular switches in the regulation of certain biological functions by switching to ordered conformation upon molecular recognition, as is the case for DNA binding, protein-protein interactions, and other events⁵⁴. In spite of their pronounced flexibility, disordered regions can adopt a fixed three-dimensional structure upon binding to other macromolecules. In the nucleus, HMGB1 binds to DNA in a non-sequence-specific manner during transcription, and this ability may also be due to the disordered region in the A-box domain⁵⁵. An experimental study has suggested that the binding sites in HMGB1 comprise a heparin-

ARTICLE

binding site (residues 2–10), an LPS binding site (residues 80–96) in the TLR4 binding domain, and a RAGE binding domain (residues 150–183)⁵⁶. *In silico* prediction also suggested that these are disordered regions. These disordered regions are typically involved in protein–protein interactions, playing crucial roles in signal transduction. Flexible linkers allow the connecting domains to twist and rotate freely, to recruit their binding partners or for those binding partners to induce larger scale inter-domain conformation changes *via* protein domain dynamics⁵⁷. The RMSD results of linker 1 loop, the TLR4 binding domain, and linker 2 loop suggest that they are especially dynamic regions in the inactive HMGB1 species, in contrast to the active HMGB1^{S-S} (Fig. S10a–c). The majority of proteins show strong correlations between structure and dynamics⁵⁸. Many disordered proteins have their binding affinity with their receptors regulated by redox state modification⁵⁹. Thus, it has been proposed that the flexibility of disordered proteins facilitates the different conformational requirements for binding to modifying enzymes, as well as their receptors. Intrinsic disorder is particularly enriched in proteins implicated in cell signaling, transcription, and chromatin remodeling functions⁵⁴.

With new insights in hand, we are poised to begin addressing the structural requirements for HMGB1 chemokine activity. It has recently been clarified that the redox state requirements for HMGB1^{SH-SH} to induce potent chemotactic activity, recruiting neutrophils and monocytes to inflammatory sites, is distinctly different than that required for cytokine-inducing activity⁹. All three cysteines must be fully reduced for HMGB1 (HMGB1^{SH-SH}) to elicit chemotactic activity. The loss of cytokine activity in HMGB1^{SH-SH} might be due to the loss of synchronized domain movements, as well as rearrangement of the active TLR4 binding domain. A possible basis for chemokine activity is the high flexibility of fully reduced HMGB1^{SH-SH}, which presents more dynamic linker loops compared to other inactive forms (HMGB1^{SO3H}, HMGB1^{C23A}, HMGB1^{C45A}, HMGB1^{C106A}) (Movies S2–S6). Porcupine plot clearly shows the difference in domain movement (Fig S11). This flexibility may provide a conformational state that is necessary for chemokine activity. Another basis for such activity may lie in the disordered regions of this protein. The free movement of linker regions provides enough

conformational freedom for certain activities. However, this high flexibility might also be the reason behind the loss of cytokine activity, due to the loss of the precise geometry required of the active site in the TLR4 binding domain loop, and in the general the shape of the B-box domain. HMGB1^{SH-SH} enables formation of a heterocomplex with the chemokine CXCL12 (stromal cell-derived factor 1), which signals via the CXCR4 receptor complex in a synergistic mode⁹. Although, it should be noted that the presence of cysteine residues is not required for chemotaxis, since they can be substituted with serine residues while preserving activity⁹. Terminal oxidation of any of the cysteines by reactive oxygen species completely abrogates the chemotactic activity². HMGB1^{SO3H} loses both cytokine and chemotactic activity. Maintaining the precise geometry of the active site in the TLR4 binding domain loop is important for cytokine activity. In the case of HMGB1^{SO3H}, the oxidation of C106 affects the domain movement of the B-box domain, leading to inactivity. A comparison of HMGB1^{SH-SH} and HMGB1^{SO3H} suggests that the oxidation of C23 and C45 affects mobility of the A-box domain. In the case of HMGB1^{SO3H}, the distance between C23 and C45 was around 4 Å, while in case of HMGB1^{SH-SH}, the distance was about 3 Å (Fig. 2c). Perhaps the mobility of the A-box domain is needed for chemotactic behavior. In case of HMGB1^{C23A}, the protein was less flexible as a whole and the mutation seemed to affect the loop regions, including the TLR4 binding domain (Movie S4). The loss in secondary structural content of the protein might result from the mutation of HMGB1^{C106A}, leading to a change in the active TLR4 binding domain loop, which in turns affects the activity. The linker displays a reduced flexibility in the mutant, which is reflected in the limited conformational space sampled by the domain. It is worth noting that changes in the redox state affect the domain movement. The most noticeable change arising from altered HMGB1 redox states was found in the TLR4 binding domain region. Importantly, our present work allows the mapping of changes in protein dynamics under different redox states. Although the time scale we used in MD simulation is not enough to fully understand the biologically important processes, we believe the following observations from our study are helpful to understand the structural changes in different states of HMGB1. The RMSD graph (Fig. 2) shows that, after 20 ns of MD simulation, all the

ARTICLE

Journal Name

HMGB1 states were equilibrated and deviated around 1 Å, except the mutated HMGB1^{C23A} and HMGB1^{C45A} states. In addition, the eigenvectors of the covariance matrix for protein atomic fluctuations also specified that the movement of proteins attained their equilibrium in the first ten eigenvectors. Though the intrinsically disorder proteins attained extremely flexible conformations in the MD simulation, the time period of equilibrated region can be considered for data analysis and subsequent atomic flexibility nature. In our study, 80% of the equilibrated region was considered and used for further analysis. Moreover, the free energy basin generated from the FEL graph also shows that each HMGB1 structure produced a maximum of four free energy minimum clusters and there was no wide range of folding mechanism. Hence, we believe that 80% of equilibrated time period is sufficient to understand the structural fluctuation and subsequent functionality.

First, we chose the lowest energy structure of active HMGB1^{S-S} from the FEL and docked it with a model of the TLR4/MD2 complex. Furthermore, we performed MD simulations and binding affinity calculations to predict the interaction between HMGB1 and key residues in the TLR4/MD2 complex. The docking results suggests that there are hydrogen bonds, ionic interactions, as well as few hydrophobic interactions, between TLR4/MD2 and HMGB1^{S-S}. Our docked model also suggests the presence of hydrophobic interactions between MD2 and HMGB1^{S-S}. This suggests that the hydrophobic interactions play a key role in the binding of the MD2/HMGB1^{S-S} complex. The TLR4 binding domain loop of HMGB1^{S-S} may interact with TLR4/MD2, according to reported studies^{2,3}. Our docking, MD simulation, and alanine scanning results have suggested important residues in HMGB1 (C23, C45, F89, K90, K96, S100, F102, F103, L104 and F105). Interacting TLR4 residues include D238, Q266, N268, D294, K146, K150, K147, K154, Q135, D238, Q266, and D294. Interacting MD2 residues include W23, S28, S45, N47, Q53, K58, and N158. The MM/PBSA results suggest that both hydrophobic and van der Waals interactions play a key role in maintaining the stability of the TLR4/MD2/HMGB1^{S-S} complex. HMGB1^{S-S} was incapable of binding directly to TLR4 in the absence of MD2, and this might be due to the hydrophobic interactions, which along with other interactions, are essential for keeping the complex

stable. The docking, MD simulation, and alanine scanning mutagenesis results suggest that both C23 and C45 are important in the interaction of the TLR4/MD2/HMGB1^{S-S} complex. This suggestion is supported by experimental studies¹². Alanine scanning mutagenesis suggests that C106 is less likely to contribute to the TLR4/MD2/HMGB1^{S-S} complex. The reason for this could be due to C106 facing away, unlike F103, L104, S107, and E108, which face towards MD2 and form interactions (Fig. 10d). It seems from our analysis that C106 doesn't bind to MD2 directly. However, our study suggests that by maintaining the precise TLR4 binding domain geometry, C106 makes a significant contribution, without which the interaction between these proteins is not feasible. The docked complex of HMGB1 has undergone conformational transitions, which is suggested when superimposed with the models of the lowest energy structures from the FEL (Fig. S7). Yang *et al.*³ reported that TLR4/MD2 is a mandatory HMGB1 receptor complex for cytokine production in macrophages, and that the interaction requires the HMGB1^{S-S} redox isoform, which binds the TLR4 co-receptor MD2 with nanomolar avidity, similar to the binding of LPS which occurs at another MD2 site. Our docked model also suggests that HMGB1^{S-S} binds to MD2, doing so at a site that differs from the LPS binding site. MD2-deficient macrophages have markedly reduced levels of HMGB1-mediated nuclear factor- κ B (NF- κ B) translocation and tumor necrosis factor (TNF) release⁵⁰. The reason for the reduced release could be that the support provided by MD2 is substantial, without which the interaction with TLR4 will be very weak. Yang *et al.* suggested that inhibition could be achieved using a novel peptide (P5779), which could bind to the hydrophobic pocket of MD2, forming maximal van der Waals interactions with the surrounding hydrophobic residues of MD2. According to Yang *et al.*, this would likely inhibit interactions with HMGB1^{S-S}⁵⁰. It is not clear from our docked model how the suggested docked peptide interactions could inhibit the interactions of HMGB1^{S-S} with TLR4/MD2. Considering the limitations in experimentally detecting the post-translational modifications of HMGB1², our study provides valuable insights into the biology of HMGB1. Moreover, the structural determination of TLRs remains difficult and time-consuming, and thus *in silico* studies are very helpful for understanding their molecular interactions. In summary, we provide a comprehensive picture of the

ARTICLE

Journal Name

structural and dynamic consequences of different redox states for HMBG1, as well as the molecular interactions of the TLR4/MD2/HMGB1^{S-S} complex. Our study may provide key insights into HMGB1 function, and may inform the development of rational therapeutics.

Journal Name

ARTICLE

Acknowledgements

This work was supported by the Mid-Career Researcher Program through the National Research Foundation of Korea, funded by the Ministry of Education, Science, and Technology (NRF-2015R1A2A2A09001059), and by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (HI14C1992). This work was also partially supported by a grant from the Priority Research Centers Program (NRF 2012-0006687).

Competing financial interests

The authors declare that they have no conflicts of interest with the contents of this article.

Author Contributions

S.P. and S.C. planned the experiments. S.P., P.D., D.Y., A.A. and H-K.K. performed the experiments and analyzed the data. S.C. contributed material. S.P. and S.C. wrote the manuscript.

Foot Notes

This article contains supplemental information.

References

1. H. E. Harris, U. Andersson and D. S. Pisetsky, *Nat Rev Rheumatol*, 2012, **8**, 195-202.
2. H. Yang, D. J. Antoine, U. Andersson and K. J. Tracey, *J Leukoc Biol*, 2013, **93**, 865-873.
3. H. Yang, H. Wang, S. S. Chavan and U. Andersson, *Mol Med*, 2015, **21**, S6-S12.
4. R. Kang, R. Chen, Q. Zhang, W. Hou, S. Wu, L. Cao, J. Huang, Y. Yu, X. G. Fan, Z. Yan, X. Sun, H. Wang, Q. Wang, A. Tsung, T. R. Billiar, H. J. Zeh, 3rd, M. T. Lotze and D. Tang, *Mol Aspects Med*, 2014, **40**, 1-116.
5. T. Ueda and M. Yoshida, *Biochim Biophys Acta*, 2010, **1799**, 114-118.
6. S. Calogero, F. Grassi, A. Aguzzi, T. Voigtlander, P. Ferrier, S. Ferrari and M. E. Bianchi, *Nat Genet*, 1999, **22**, 276-280.
7. U. Andersson and K. J. Tracey, *Annu Rev Immunol*, 2011, **29**, 139-162.
8. H. Yanai, T. Ban, Z. Wang, M. K. Choi, T. Kawamura, H. Negishi, M. Nakasato, Y. Lu, S. Hangai, R. Koshiba, D. Savitsky, L. Ronfani, S. Akira, M. E. Bianchi, K. Honda, T. Tamura, T. Kodama and T. Taniguchi, *Nature*, 2009, **462**, 99-103.
9. E. Venereau, M. Casalgrandi, M. Schiraldi, D. J. Antoine, A. Cattaneo, F. De Marchis, J. Liu, A. Antonelli, A. Preti, L. Raeli, S. S. Shams, H. Yang, L. Varani, U. Andersson, K. J. Tracey, A. Bach, M. Uguccioni and M. E. Bianchi, *J Exp Med*, 2012, **209**, 1519-1528.
10. J. Li, R. Kokkola, S. Tabibzadeh, R. Yang, M. Ochani, X. Qiang, H. E. Harris, C. J. Czura, H. Wang, L. Ulloa, H. Wang, H. S. Warren, L. L. Moldawer, M. P. Fink, U. Andersson, K. J. Tracey and H. Yang, *Mol Med*, 2003, **9**, 37-45.
11. M. Stros, *Biochim Biophys Acta*, 2010, **1799**, 101-113.
12. H. Yang, H. S. Hreggvidsdottir, K. Palmblad, H. Wang, M. Ochani, J. Li, B. Lu, S. Chavan, M. Rosas-Ballina, Y. Al-Abed, S. Akira, A. Bierhaus, H. Erlandsson-Harris, U. Andersson and K. J. Tracey, *Proc Natl Acad Sci U S A*, 2010, **107**, 11942-11947.
13. A. Roy, A. Kucukural and Y. Zhang, *Nat Protoc*, 2010, **5**, 725-738.
14. J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang, *Nat Methods*, 2015, **12**, 7-8.
15. N. Eswar, D. Eramian, B. Webb, M. Y. Shen and A. Sali, *Methods Mol Biol*, 2008, **426**, 145-159.
16. P. Suresh and S. Choi, Structural models of active and inactive redox states structures of HMGB1, https://figshare.com/articles/Structural_models_of_active_and_inactive_structure_of_human_High_mobility_group_box_1_HMGB1/_3580989, (accessed 26th August, 2016).
17. S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845-854.
18. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950-1958.
19. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys*, 1983, **79**, 926-935.
20. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Chem. Phys*, 1997, **18**, 1463-1472.
21. L. Verlet, *Physical Review*, 1967, **159**, 98-103.
22. T. Darden, D. York and L. Pedersen, *J. Chem. Phys*, 1993, **98**, 10089-10092.
23. S. S. Negi, C. H. Schein, N. Oezguen, T. D. Power and W. Braun, *Bioinformatics*, 2007, **23**, 3397-3399.
24. B. S. Park, D. H. Song, H. M. Kim, B. S. Choi, H. Lee and J. O. Lee, *Nature*, 2009, **458**, 1191-1195.
25. C. Dominguez, R. Boelens and A. M. Bonvin, *J Am Chem Soc*, 2003, **125**, 1731-1737.
26. S. J. de Vries, M. van Dijk and A. M. Bonvin, *Nat Protoc*, 2010, **5**, 883-897.
27. A. Amadei, A. B. Linssen and H. J. Berendsen, *Proteins*, 1993, **17**, 412-425.
28. C. C. David and D. J. Jacobs, *Methods Mol Biol*, 2014, **1084**, 193-226.
29. L. Skjaerven, X. Q. Yao, G. Scarabelli and B. J. Grant, *BMC Bioinform.*, 2014, **15**, 399.
30. D. Spiliotopoulos, A. Spitaleri and G. Musco, *PLoS One*, 2012, **7**, e46902.
31. I. Massova and P. A. Kollman, *Journal of the American Chemical Society*, 1999, **121**, 8133-8143.
32. N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc Natl Acad Sci U S A*, 2001, **98**, 10037-10041.
33. C. Paissoni, D. Spiliotopoulos, G. Musco and A. Spitaleri, *Comput. Phys. Commun.*, 2014, **185**, 2920-2929.

34. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605-1612.
35. H. H. Kim, S. J. Park, J. H. Han, C. Pathak, H. K. Cheong and B. J. Lee, *Biochim Biophys Acta*, 2015, **1854**, 449-459.
36. R. Sanchez-Giraldo, F. J. Acosta-Reyes, C. S. Malarkey, N. Saperas, M. E. Churchill and J. L. Campos, *Acta Crystallogr D Biol Crystallogr*, 2015, **71**, 1423-1432.
37. J. Wang, N. Tochio, A. Takeuchi, J. I. Uewaki, N. Kobayashi and S. I. Tate, *Biochem Biophys Res Commun*, 2013, **441**, 701-702.
38. J. P. Rowell, K. L. Simpson, K. Stott, M. Watson and J. O. Thomas, *Structure*, 2012, **20**, 2014-2024.
39. K. Stott, M. Watson, M. J. Bostock, S. A. Mortensen, A. Travers, K. D. Grasser and J. O. Thomas, *J Biol Chem*, 2014, **289**, 29817-29826.
40. L. P. Kozlowski and J. M. Bujnicki, *BMC Bioinform.*, 2012, **13**, 111.
41. M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo and A. Sali, *Annu Rev Biophys Biomol Struct*, 2000, **29**, 291-325.
42. D. Taylor, G. Cawley and S. Hayward, *PLoS One*, 2013, **8**, e81224.
43. R. B. Best, *J Phys Chem B*, 2013, **117**, 13235-13244.
44. J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes, *Annu Rev Phys Chem*, 1997, **48**, 545-600.
45. A. Gershenson, L. M. Gerasch, A. Pastore and S. E. Radford, *Nat Chem Biol*, 2014, **10**, 884-891.
46. A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. Iben, T. B. Sauke, E. Shyamsunder and R. D. Young, *Proc Natl Acad Sci U S A*, 1985, **82**, 5000-5004.
47. L. Q. Yang, P. Sang, Y. Tao, Y. X. Fu, K. Q. Zhang, Y. H. Xie and S. Q. Liu, *J Biomol Struct Dyn*, 2014, **32**, 372-393.
48. D. Prada-Gracia, J. Gomez-Gardenes, P. Echenique and F. Falo, *PLoS Comput Biol*, 2009, **5**, e1000415.
49. J. N. Onuchic and P. G. Wolynes, *Curr Opin Struct Biol*, 2004, **14**, 70-75.
50. H. Yang, H. Wang, Z. Ju, A. A. Ragab, P. Lundback, W. Long, S. I. Valdes-Ferrer, M. He, J. P. Pribis, J. Li, B. Lu, D. Gero, C. Szabo, D. J. Antoine, H. E. Harris, D. T. Golenbock, J. Meng, J. Roth, S. S. Chavan, U. Andersson, T. R. Billiar, K. J. Tracey and Y. Al-Abed, *J Exp Med*, 2015, **212**, 5-14.
51. H. Yang, M. Ochani, J. Li, X. Qiang, M. Tanovic, H. E. Harris, S. M. Susarla, L. Ulloa, H. Wang, R. DiRaimo, C. J. Czura, H. Wang, J. Roth, H. S. Warren, M. P. Fink, M. J. Fenton, U. Andersson and K. J. Tracey, *Proc Natl Acad Sci U S A*, 2004, **101**, 296-301.
52. S. A. Lee, M. S. Kwak, S. Kim and J. S. Shin, *Yonsei Med J*, 2014, **55**, 1165-1176.
53. X. Y. Meng, B. Li, S. Liu, H. Kang, L. Zhao and R. Zhou, *Sci Rep*, 2016, **6**, 22128.
54. R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, *Chem Rev*, 2014, **114**, 6589-6631.
55. K. Mitsouras, B. Wong, C. Arayata, R. C. Johnson and M. Carey, *Mol Cell Biol*, 2002, **22**, 4390-4401.
56. E. Ranzato, S. Martinotti, M. Pedrazzi and M. Patrone, *Cells*, 2012, **1**, 699-710.
57. V. P. Reddy Chichili, V. Kumar and J. Sivaraman, *Protein Sci*, 2013, **22**, 153-167.
58. U. Hensen, T. Meyer, J. Haas, R. Rex, G. Vriend and H. Grubmuller, *PLoS One*, 2012, **7**, e33931.
59. D. Reichmann and U. Jakob, *Curr Opin Struct Biol*, 2013, **23**, 436-442.

Figure legends

Fig. 1 | Modeled structure of HMGB1. **(a)** The figure shows the lowest energy structure of modeled full-length active HMGB1. Protein is shown in ribbon representation. The A-box domain consists of residues 1–79 (yellow). Linker loop 1 consists of residues 80–88 (green). The B-box domain consists of residues 89–162 (blue). The TLR4 binding domain consists of residues 89–108 (red). Linker loop 2 consists of residues 89–162 (pink). The C-terminal tail consists of residues 186–215 (orange). The figure was generated using PyMOL. **(b)** Predictions of disordered regions in HMGB1. The predictions were made using MetaServer. Each color denotes the primary method of prediction. **(c)** Isoforms of HMGB1, their MD2 binding ability, and their cytokine/chemokine activity are shown.

Fig. 2 | Stability of the redox states of HMGB1. Color codes: HMGB1^{S-S} (green), HMGB1^{SH-SH} (red), HMGB1^{SO3H} (yellow), HMGB1^{C23A} (orange), HMGB1^{C45A} (gray) and HMGB1^{C106A} (blue). Molecular dynamics simulations (100 ns) of the different redox states of HMGB1. Prepared using Matplotlib. **(a)** Graph of RMSD of the protein backbone atoms with respect to the initial structure. **(b)** RMSF values for each of the simulated complexes are indicated and compared. **(c)** Radius of gyration (R_g) values. **(d)** Minimum distance between C23 and C45 residues.

Fig. 3 | Free energy landscape (FEL) of the HMGB1^{S-S} model. The FEL was calculated using the first two principal components as reaction coordinates for HMGB1^{S-S}. The major basins are labeled (I to III). Representative lowest energy structures from the three states, are shown. The TLR4 binding domain is identified by arrows. Dark blue represents the most favorable conformation.

Fig. 4 | Free energy landscape (FEL) of the HMGB1^{SH-SH} model. The FEL was calculated using the first two principal components as reaction coordinates for HMGB1^{SH-SH}. The major basins are labeled (I to IV). Representative lowest energy structures from the four states are shown. The transition of the structure is clearly visible. Dark blue represents the most favorable conformation.

Fig. 5 | Free energy landscape (FEL) of the HMGB1^{SO3H} model. The FEL was calculated using the first two principal components as reaction coordinates for HMGB1^{SO3H}. The major basins are labeled (I to III). Representative structures from the three states are shown. Dark blue represents the most favorable conformation.

Fig. 6 | Free energy landscape (FEL) of HMGB1^{C23A} model. The FEL was calculated using the first two principal components as reaction coordinates for the HMGB1^{C23A} mutant. The major basins are labeled (I to III). Representative structures from the three states are shown. Dark blue represents the most favorable conformation.

Fig. 7 | Free energy landscape (FEL) of HMGB1^{C45A} model. The FEL was calculated using the first two principal components as reaction coordinates for HMGB1^{C45A} mutant. The major basins are labeled (I and II). Representative structures from the three states are shown. Dark blue represents the most favorable conformation.

Fig. 8 | Free energy landscape (FEL) of the HMGB1^{C106A} model. The FEL was calculated using the first two principal components as reaction coordinates for the HMGB1^{C106A} mutant. The major basins are labeled (I to IV). Representative structures from the three states are shown. Dark blue represents the most favorable conformation.

Fig. 9 | Binding poses of active and inactive HMGB1 redox states. (a) HMGB1^{S-S}, **(b)** HMGB1^{SH-SH}, **(c)** HMGB1^{SO3H}, **(d)** HMGB1^{C23A}, **(e)** HMGB1^{C45A}, and **(f)** HMGB1^{C106A}.

Fig. 10 | The binding mode of HMGB1^{S-S} docked with the TLR4/MD2 complex. (a) The overall binding mode of TLR4/MD2/HMGB1^{S-S} is shown in cartoon representation. The TLR4 domain is shown in blue and cyan. MD2 is shown in green and yellow. The docked HMGB1^{S-S} model is shown in red. **(b)** Interacting residues of TLR4/MD2 are shown in red, with residue labels. **(c)** Interacting residues of HMGB1^{S-S} are shown in yellow, with residue labels. **(d)** The TLR4/MD2/HMGB1^{S-S} complex is shown. For clarity, only a few important hydrogen bonding residues are depicted.

ARTICLE

Journal Name

Table 1. Residues in the TLR4/MD2 interaction with HMGB1^{S-S}, within a distance of 3.5 Å. An asterisk (*) denotes experimentally confirmed active residues.

Complex	Hydrogen bond interacting residues	Hydrophobic interacting residues	Ionic interactions
MD2/HMGB1	S28.OG ... K96.NZ* S45.OG ... R110.NZ N47.ND2 ... S107.OG E53.OE1 ... K96.NZ* K58.O ... Y155.OH N158.OCZ ... Y173.NZ	W23.CZ3 ... F103.CZ*	E53.OE1 ... K96.NZ* N158.OCZ ... K173.NZ
TLR4/HMGB1	D238.OD2 ... K146.NZ D238.OD1 ... K150.NZ E266.OE2 ... K147.NZ N268.O ... K150.NZ D294.OD2 ... K154.NZ		E135.OE2 ... K112.NZ* D238.OD1 ... K146.NZ D238.OD2 ... K150.NZ E266.OE2 ... K147.NZ D294.OD2 ... K154.NZ

Table 2. MM/PBSA binding free energies (kJ/mol) for the TLR4 binding domain (residues 89–108) of active HMGB1^{S-S} and the alanine mutants. ¹ Binding free energies; ² Polar terms; ³ Non-polar terms; ⁴ Coulombic terms; ⁵ van der Waals terms; ⁶ Polar solvation terms; ⁷ Non-polar solvation terms.

Name	ΔG^1	Δ_{polar}^2	Δ_{npolar}^3	Δ_{coul}^4	Δ_{vdW}^5	Δ_{psol}^6	Δ_{npsol}^7
TLR4-MD2-HMGB1	-117.3 +/- 40.1	454.3	-576.6	1407.8 +/- 100.1	-512.4 +/- 7.4	-953.5 +/- 97.7	-59.2 +/- 1.0
C23A	-19.1 +/- 38.7	516.6	-535.7	1483.5 +/- 95.1	-476.6 +/- 7.7	967.0 +/- 93.3	-59.1 +/- 1.0
C45A	25.1 +/- 37.1	531.7	-506.6	1494.3 +/- 95.0	-447.4 +/- 9.7	962.7 +/- 93.2	-59.2 +/- 1.0
F89A	-105.4 +/- 38.2	433.8	-539.2	1415.9 +/- 95.0	-481.6 +/- 7.2	982.1 +/- 93.8	-57.6 +/- 1.0
K90A	-30.8 +/- 41.7	542.1	-573.0	2318.1 +/- 100.1	-512.1 +/- 7.5	-1775.9 +/- 100.1	-60.8 +/- 1.0
D91A	242.0 +/- 37.2	803.9	-561.9	878.8 +/- 101.7	-502.5 +/- 7.1	75.0 +/- 94.7	-59.4 +/- 1.0
N93A	150.1 +/- 39.0	693.1	-543.1	1660.2 +/- 96.7	-484.1 +/- 7.5	-967.1 +/- 97.7	-59.0 +/- 1.0
K96A	-113.9 +/- 41.6	438.8	-552.6	2278.5 +/- 93.1	-494.5 +/- 7.7	-1839.8 +/- 95.3	-58.1 +/- 1.1
R97A	498.9 +/- 43.1	1028.5	-529.6	2987.4 +/- 97.7	-472.1 +/- 8.6	-1958.9 +/- 95.4	-57.5 +/- 1.1
S100A	-98.0 +/- 40.6	466.2	-564.1	1425 +/- 98.8	-505.0 +/- 7.6	-959.3 +/- 97.7	-59 +/- 1.1
F102A	-93.6 +/- 39.9	455.3	-548.8	1396.6 +/- 100.1	-490.1 +/- 7.7	941.4 +/- 97.7	-58.8 +/- 1.0
F103A	-93.1 +/- 39.7	448.8	-541.8	1400.7 +/- 100.3	-483.9 +/- 7.4	951.9 +/- 98.2	-57.9 +/- 1.0
L104A	-47.6 +/- 38.4	489.1	-536.7	1416.4 +/- 99.5	-478.1 +/- 7.0	927.3 +/- 95.6	-58.6 +/- 1.0
F105A	-87.2 +/- 39.1	439.1	-526.3	1401.8 +/- 95.7	-468.1 +/- 6.3	962.7 +/- 94.0	-58.2 +/- 1.0
C106A	-141.4 +/- 39.9	418.8	-560.2	1377.9 +/- 99.7	-501.1 +/- 7.5	-959.1 +/- 97.6	-59.1 +/- 1.0
S107A	-91.4 +/- 40.1	474.2	-565.5	1452.6 +/- 94.2	-506.4 +/- 7.6	978.5 +/- 93.4	-59.1 +/- 1.0
E108A	142.6 +/- 37.3	701.9	-559.2	649.8 +/- 99.0	-500.2 +/- 7.4	52.1 +/- 95.0	-59.0 +/- 1.0

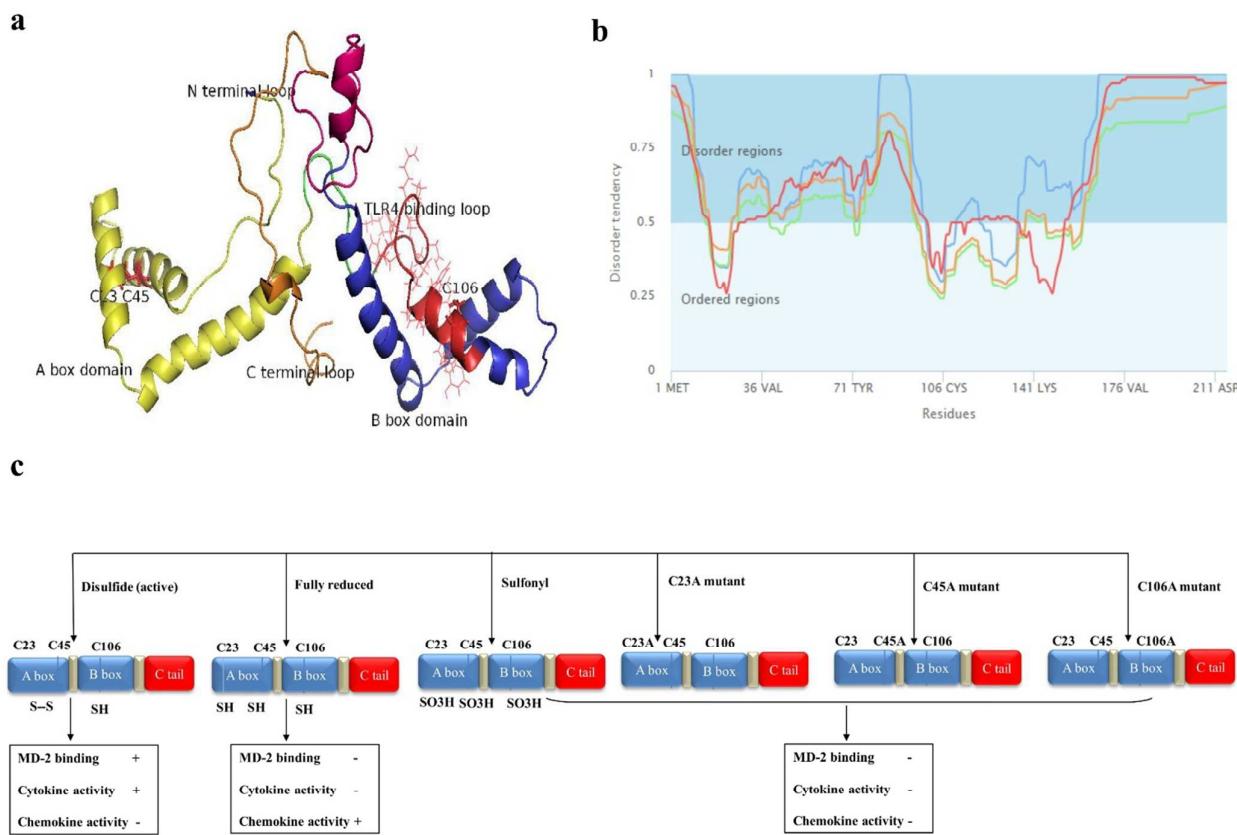
Fig. 1

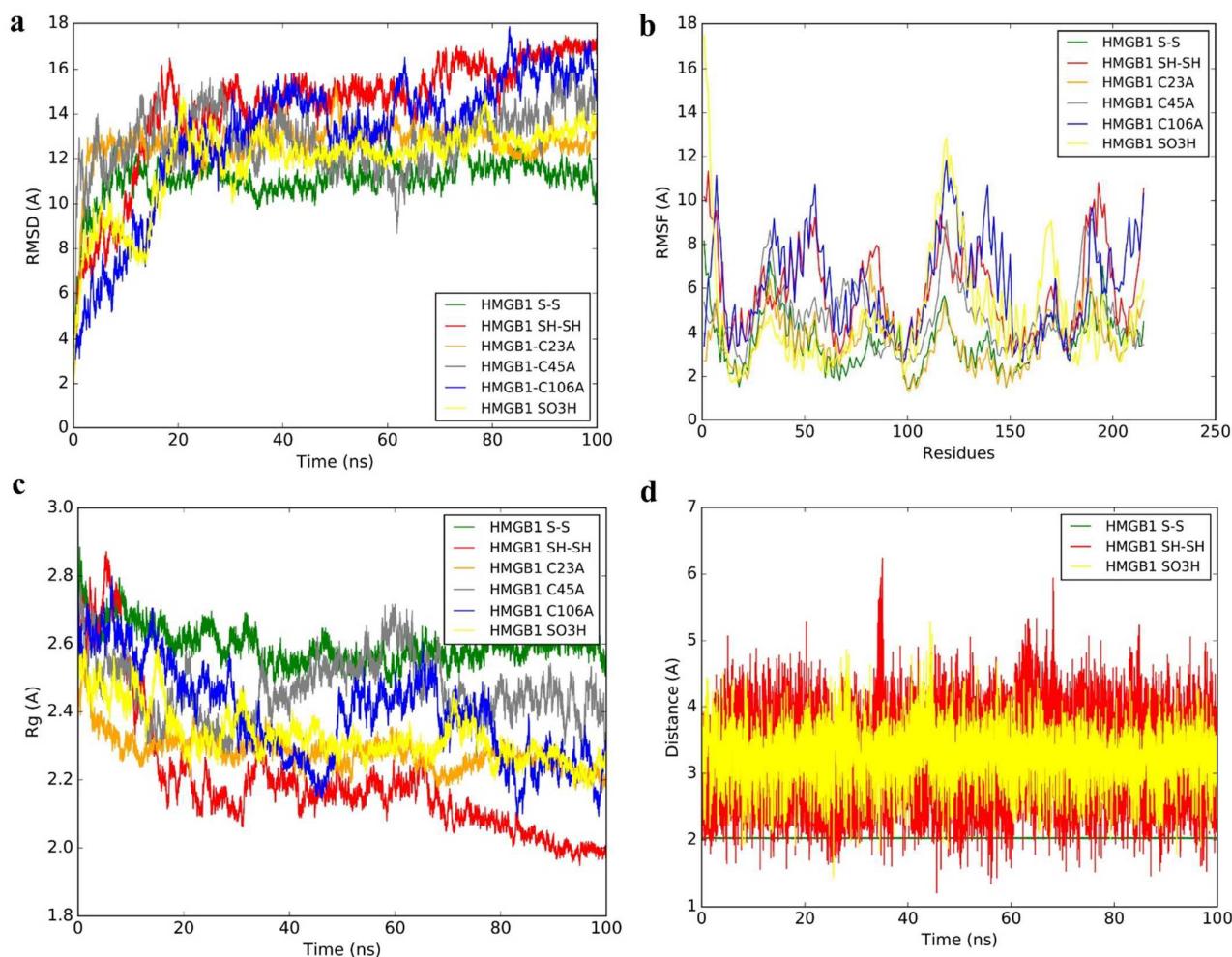
Fig. 2

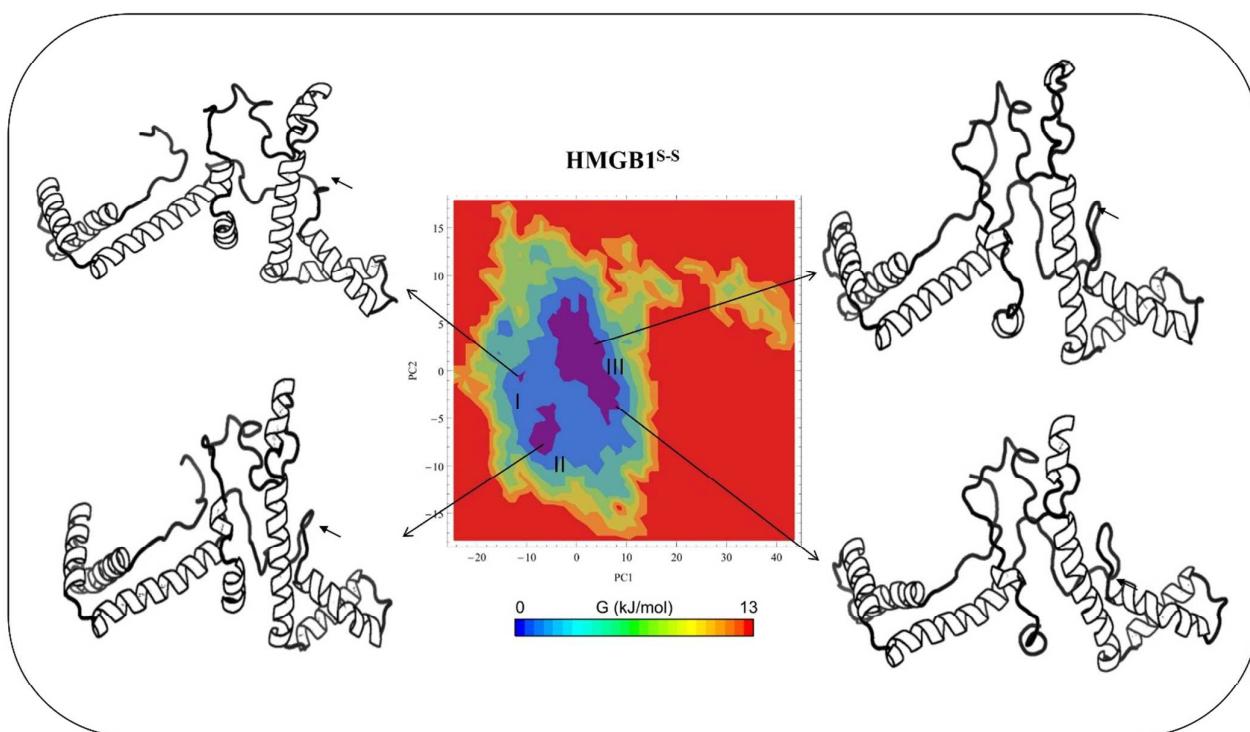
Fig. 3

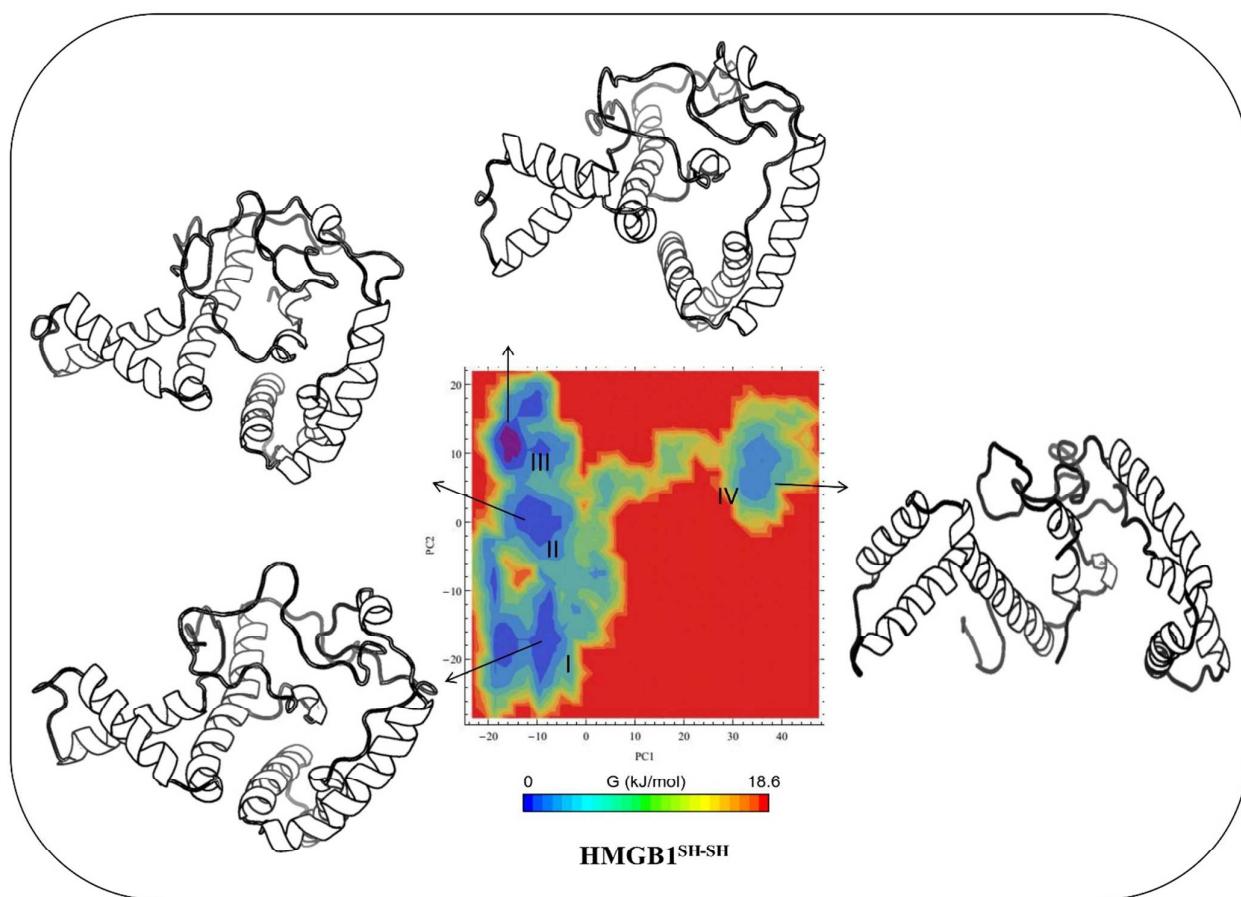
Fig. 4

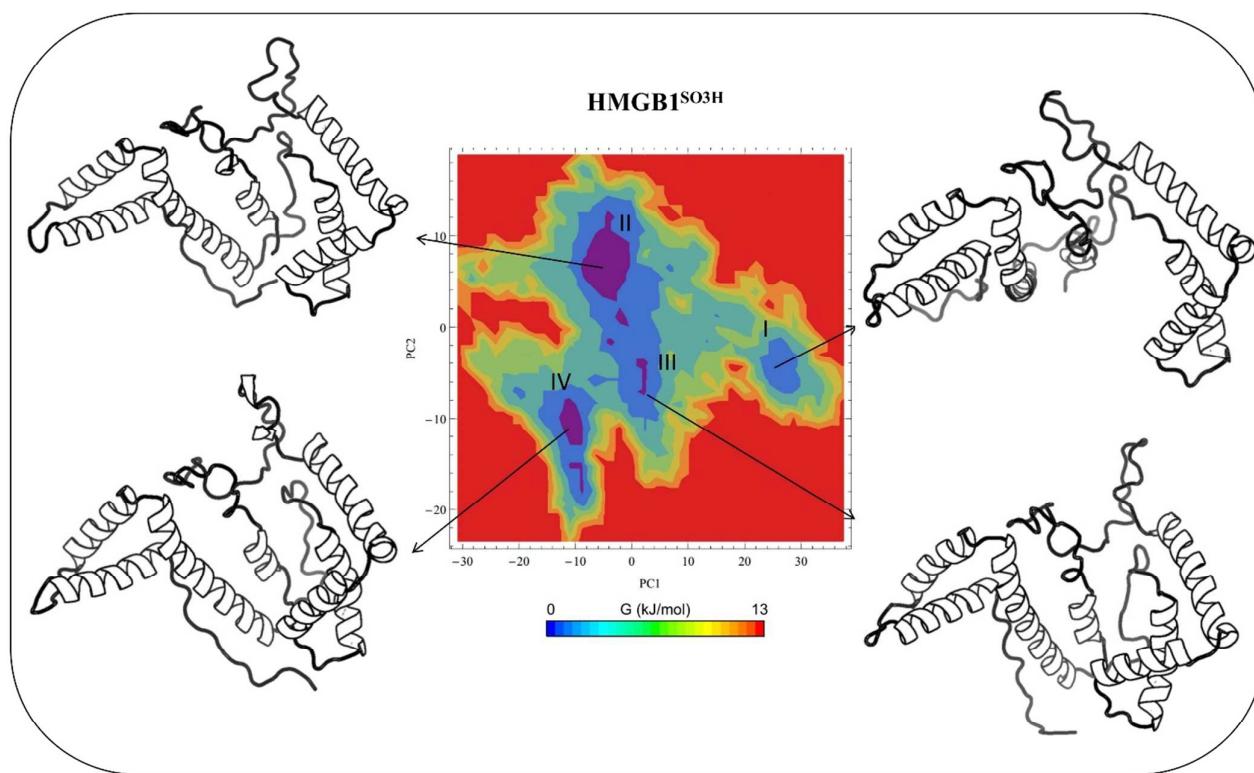
Fig. 5

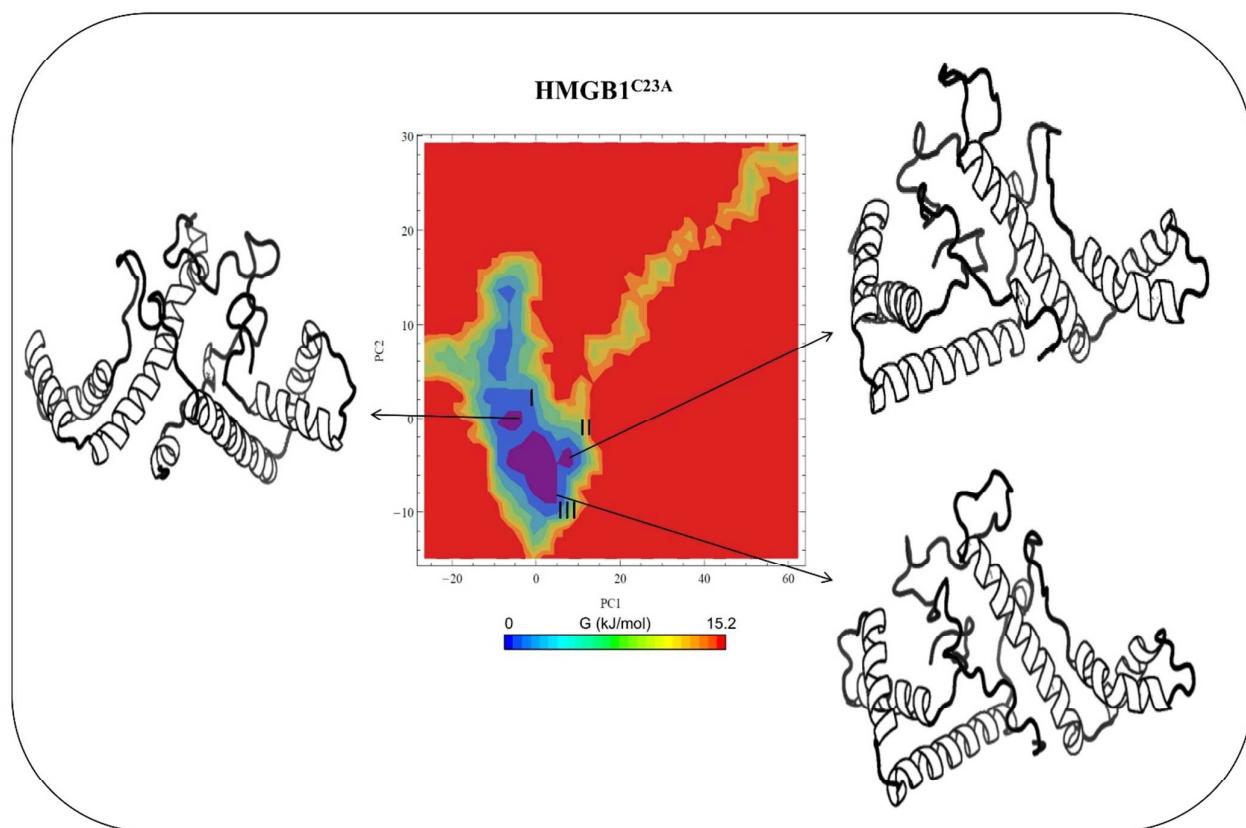
Fig. 6

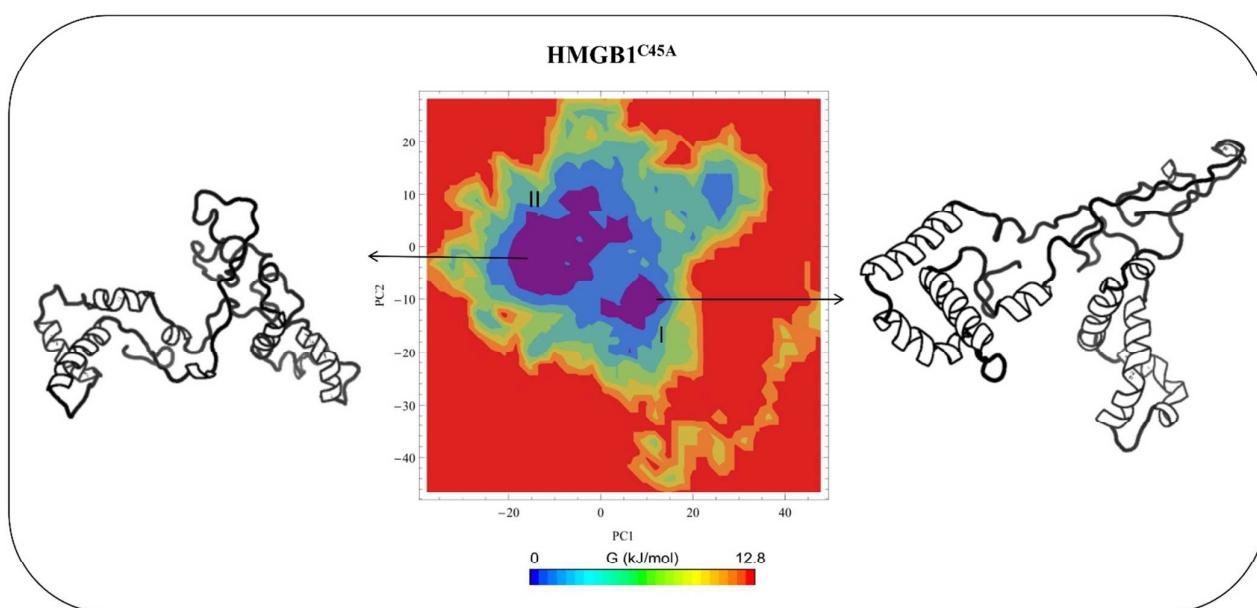
Fig. 7

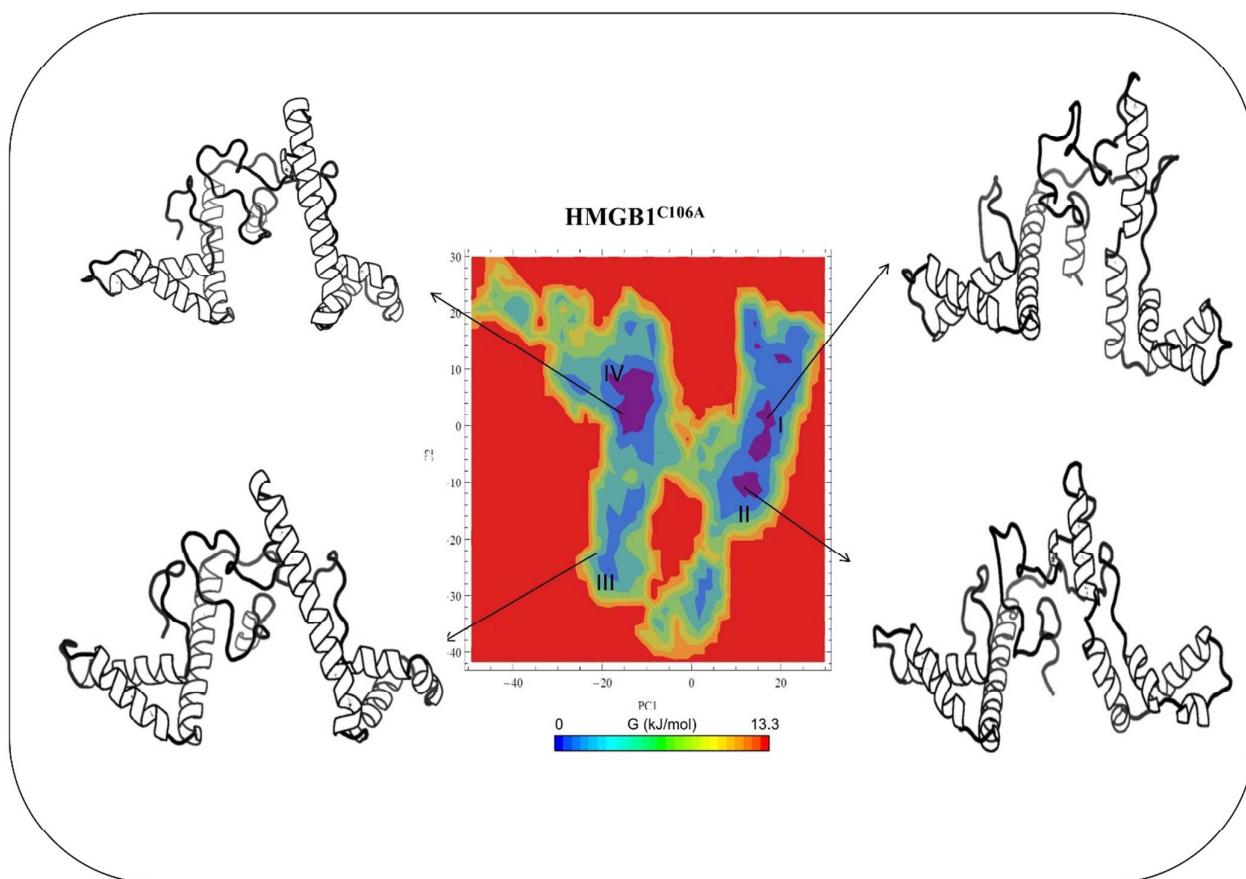
Fig. 8

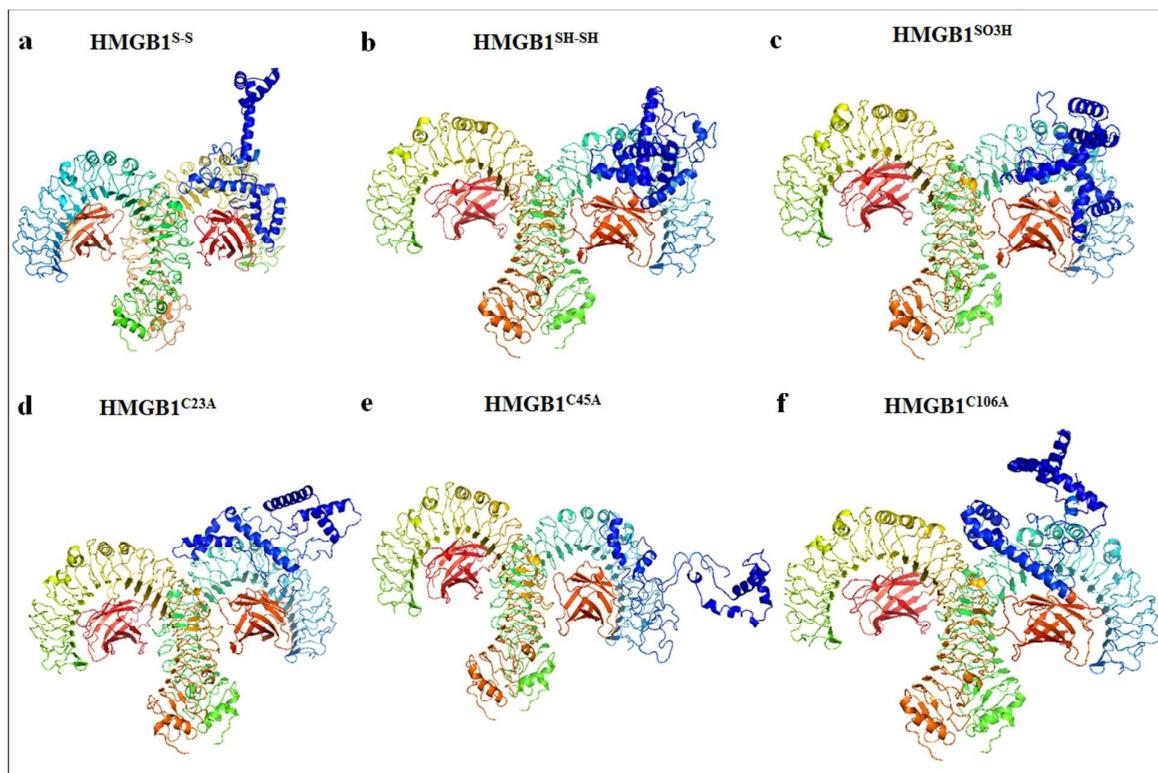
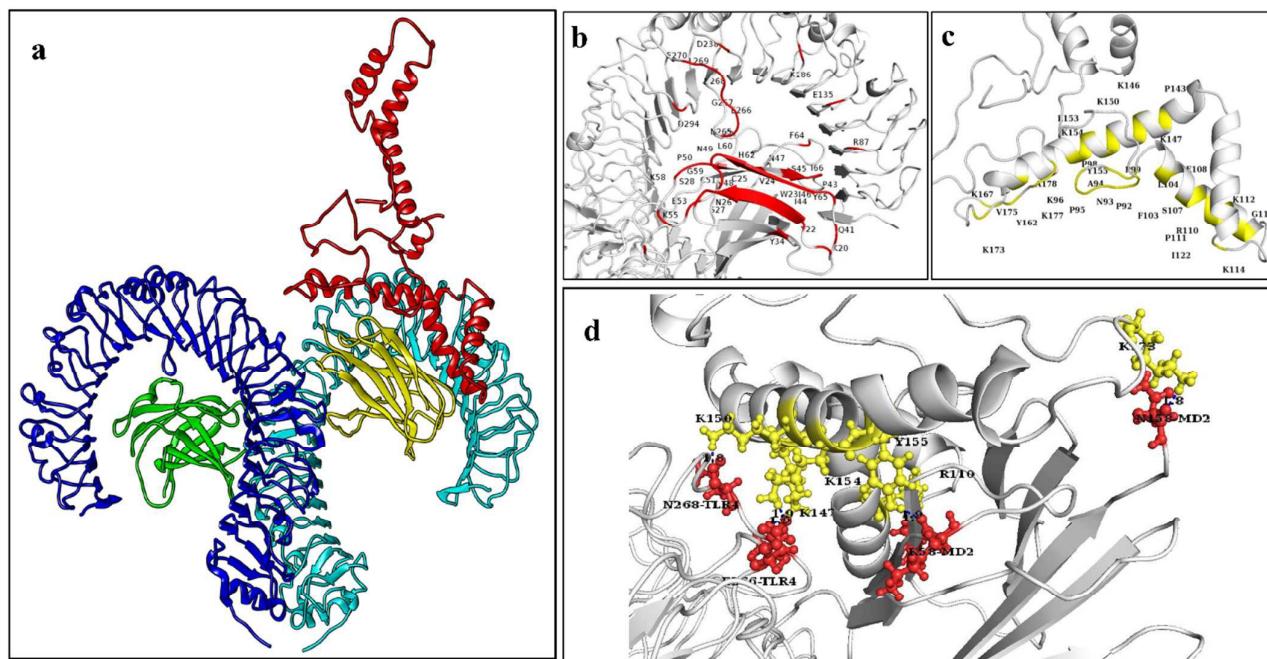
Fig. 9

Fig. 10

Graphic Figure (Fig. 3)

Highlights of the work: We have modelled and simulated different states of HMGB1, suggesting that the fully reduced HMGB1 maintains the inter-domain movements during the activity.

