JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Identification of a New Binding Site in *E. coli* FabH using Molecular Dynamics Simulations: Validation by Computational Alanine Mutagenesis and Docking Studies

Divya Ramamoorthy,[†] Edward Turos,[†,§,∥] and Wayne C. Guida[†,‡,§,∥,*]

[†]Department of Chemistry, University of South Florida, 4202 E. Fowler Avenue, Tampa, Florida 33620, United States

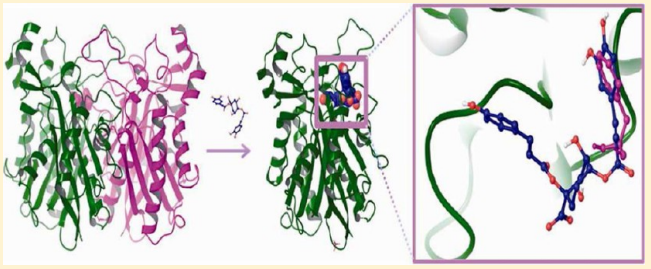
[‡]Drug Discovery Department, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Drive, Tampa, Florida 33612, United States

[§]Center for Molecular Diversity in Drug Design, Discovery and Delivery, 4202 E. Fowler Avenue, Tampa, Florida 33620, United States

[∥]Center for Drug Discovery and Innovation, 4202 E. Fowler Avenue, Tampa, Florida 33620, United States

Supporting Information

ABSTRACT: FabH (Fatty acid biosynthesis, enzyme H, also referred to as β-ketoacyl-ACP-synthase III) is a key condensing enzyme in the type II fatty acid synthesis (FAS) system. The FAS pathway in bacteria is essential for growth and survival and vastly differs from the human FAS pathway. Enzymes involved in this pathway have arisen as promising biomolecular targets for discovery of new antibacterial drugs. However, currently there are no clinical drugs that selectively target FabH, and known inhibitors of FabH all act within the active site. FabH exerts its catalytic function as a dimer, which could potentially be exploited in developing new strategies for inhibitor design. The aim of this study was to elucidate structural details of the dimer interface region by means of computational modeling, including molecular dynamics (MD) simulations, in order to derive information for the structure-based design of new FabH inhibitors. The dimer interface region was analyzed by MD simulations, trajectory snapshots were collected for further analyses, and docking studies were performed with potential small molecule disruptors. Alanine mutation and docking studies strongly suggest that the dimer interface could be a potential target for anti-infection drug discovery.



INTRODUCTION

Infectious diseases cause about 13 million fatalities per year, accounting for 25% of deaths worldwide.[1] A repercussion of excessive usage of antibiotics has been the development of resistance to almost every class of antimicrobials. Innovative techniques are required for development of novel antibiotics to overcome the problem of resistance against currently available drugs.

Fatty acid biosynthesis has proven to be an alluring target for antibacterials. Fatty acid synthase, type II (FAS II) governs the synthesis of straight chain saturated and unsaturated fatty acids in archaeabacteria and eubacteria, while FAS I is employed by higher organisms for fatty acid biosynthesis. FAS II, which is generally utilized by bacteria, is comprised of a more dissociated system in which nine separate polypeptides are involved that may be dimeric, tetrameric, or other higher order structures.[2] A very important feature of FAS II is that the intermediates are transported through the cytoplasm by an acidic and highly conserved acyl carrier protein (ACP). FAS I possesses a similar carrier arm, but it is attached to the main polypeptide chain, which makes the intermediates in this pathway much less accessible to the other enzymes. FAS I and FAS II differ from one another and lack appreciable homology, but their mechanisms of elongation and reduction are similar. Each of the discrete enzymes in FAS II can be compared to an equivalent domain in FAS I. The significant differences in structural organization and role played by the fatty acid biosynthetic enzymes between bacteria and humans make FAS II an opportune target for antibacterial drug discovery.[3]

FabB, FabF, and FabH (Fatty acid biosynthesis, enzyme H) are the condensing enzymes in FAS II that play an important role in the initiation and elongation of fatty acid chains. FabB and FabF use acyl-ACPs as primers, while FabH uses acetyl-CoA as an acyl donor, which makes it unique among the FAS II condensing enymes.[4] FabH, also referred to as β-ketoacyl-ACP-synthase III, is a key condensing enzyme of the type II FAS system and is essential for the growth and survival of both Gram-positive and Gram-negative bacteria. It has been extensively studied in *E. coli*, although studies have also been done in *S. aureus*[5] and *M. tuberculosis*.[6] FabH is ubiquitously expressed in these organisms, indicating that the enzyme has an

Received: July 27, 2012

   dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

important regulatory function and is essential for organismal survival.[7] Some small molecules have also been shown to inhibit FabH,[8,9] which supports the concept that FabH could be a good biomolecular target. FabH does not have any close homologue in humans, and the previously mentioned architectural differences between human FAS I and bacterial FAS II make FabH an ideal candidate for the identification of novel[10] selective antibacterial drugs.

Cerulenin and thiolactomycin (Figure 1a,b) are two inhibitors that were discovered a couple of decades ago that target the condensing enzymes in the FAS pathway, but unfortunately, they exhibited poor antibacterial activities.[9] Platencin (Figure 1c) is a dual inhibitor targeting FabH and FabF and exhibits IC_{50} values of 9.17 μM and 4.58 μM toward *S. aureus* FabH and FabF, respectively.[8] Other inhibitors targeting FAS II condensing enzymes identified in the past[11−13] either failed to reach their intracellular targets or lacked target specificity.[9] These inhibitors all act within the active site, and the mode of action has been extensively studied to develop a model for active-site inhibitor design.[14]

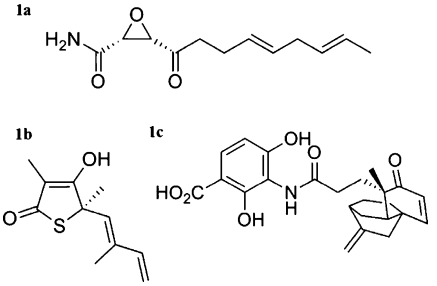


Figure 1. (a) Structure of cerulenin, a known inhibitor of FabF/FabB. (b) Structure of thiolactomycin, a known inhibitor that targets FabH and FabF/FabB. (c) Structure of platencin, a dual inhibitor for FabH and FabF.

FabH functions as a homodimer, with a monomeric molecular weight of 35 kDa,[15,16] and it is highly conserved among the key pathogens. FabH catalyzes the Claisen condensation reaction of malonyl-ACP with an enzyme-bound acetate unit (obtained from the reaction between acetyl-CoA and the catalytic residue, cysteine, with release of CoA).[6] The reaction product formed from the FabH initiated reaction, undergoes chain elongation reactions in subsequent cycles catalyzed by FabF and FabB to produce long-chain acyl-ACP products.[6] FabH, the initiation condensing enzyme of the FAS II pathway, is structurally distinct from the elongation condensing enzymes, FabF and FabB.

As shown crystallographically for small molecules co-crystallized with FabH, the dimer interface buries about 3200 Å^2 of surface area and plays an important role in substrate recognition.[17,18] Dimer formation is important for proper functioning of FabH.[6] The interface is primarily formed by residues in four loops: 84−86, 146−157, 185−217, 305−307; all of which are involved in the formation of the active site.[6] Residues that are important for substrate recognition include Trp32, Arg151, and Phe87 of the other monomer. Trp32 and Arg151 lie at the entrance to the active site[17] (Figure 2). Residue Phe87 of one monomer projects into the binding pocket of the other monomer and engages in van der Waals

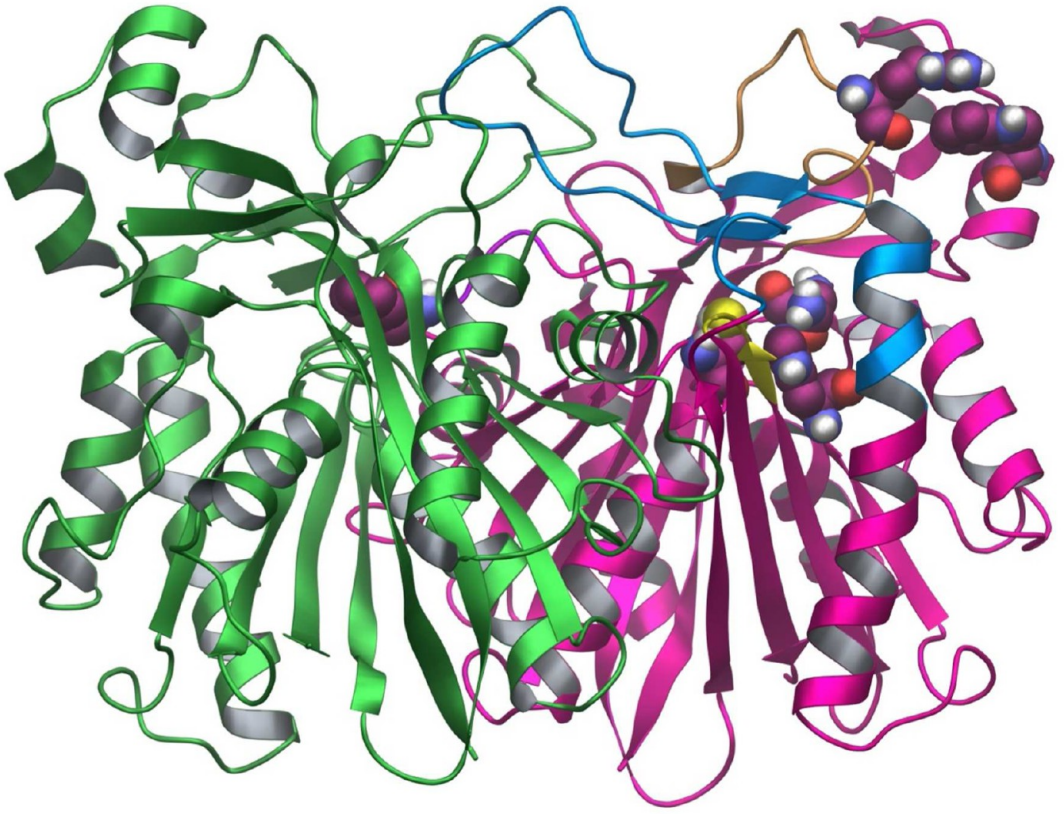


Figure 2. Crystal structure of FabH (1HN9) shown as ribbon representation. Important residues in the dimer interface region and active site are represented in cpk: Phe87 (left); Trp32 and 151 (upper right); Cys112, His244, and Asn274 (lower right).

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

interactions with acetylated Cys112 and substrate, making it crucial in determining primer specificity.[6,17] The crystal structure of FabH in the apo form has also been solved[17,18] and is reported to be disordered in the four loops that form the dimer interface. An important observation reveals that Phe87 swings away from its ligand-bound position, and the side chain of Phe87 is disordered in the tetragonal form of apo FabH.[18] Moreover, Phe87 is conserved in FabH in a number of bacterial species (see Supporting Information (SI) for sequence alignment). These observations suggest that residue Phe87 has a significant role in stabilizing the active form of the dimer.

The dissociation constant (K_D) for the monomer−dimer equilibrium in FabH was observed to be 1.00 ± 0.03 μM from ultracentrifugation studies and analytical gel filtration data,[18−20] which indicates relatively weak dimerization compared to the dissociated monomers.

In recent years, structure-based molecular modeling has emerged as an important tool in drug discovery. Computational docking including virtual screening, along with molecular dynamics simulations are widely used tools in structure-based drug design. For structure-based drug design of dimerization disruptors, a thorough knowledge of the dimer interface is essential. Sharing about 3200 Å^2 of surface area, the dimer interface is indeed a protein−protein interaction interface with each monomer having a huge dimeric partner. Molecular dynamics and docking studies have been performed previously on FabH, but they specifically targeted the biological active site.[21] Although the dimer interface has been studied in FabH, very little is known about the conformational properties of the solvated monomers.

Previously, it has been reported that in order to disrupt a protein−protein interaction, it is not necessary to target the entire interface region; it is only necessary to identify "hot spots" at the interface region that provide pockets for just a few residues that collectively make an overwhelming contribution to the binding free energy.[22] It is also known that small molecules or peptides that bind to one partner almost always compete with the natural protein partner to bind at the hot spot(s). This implies that tight binding using small molecules and or peptides can be achieved although the surface area covered by them would not be as large as the protein partner. In support of this hypothesis, several small molecule protein−protein disruptors have been identified in the past decade.[22]

Targeting the dimer interface in FabH would potentially allow the discovery of an entirely new class of antibacterial agents. Moreover, as has been previously suggested for dimerization inhibitors of HIV protease[23] that dimerization disruptors of FabH may be less susceptible to antimicrobial resistance than inhibitors that bind to the catalytic site of the enzyme. We note that it has been observed that most of the residues in the dimer interface region of FabH are conserved[24]- (see SI for sequence alignment). Although one might argue that in the absence of a disruptor there is little selective pressure for these residues to undergo mutation, it is interesting to note that in HIV protease it has been observed that mutations that are known to confer resistance to inhibitors acting at the active site occur as natural polymorphisms in HIV strains taken from patients who have never been treated with a protease inhibitor.[25] On the other hand, we recognize that resistance mechanisms in bacteria are far more complex than the ones observed in viruses like HIV, and we thank one of the reviewers for pointing out that in HIV the major source of acquired resistance is due to mutations that are a consequence of the low fidelity of the HIV-RT. Nonetheless, potent selective dimerization disruptors of FabH would allow one to test the hypothesis that such inhibitors would be less susceptible to mutational resistance.

Here, we present molecular dynamics simulations targeted at understanding the structure and conformational mobility of the solvated monomers compared to the dimer in solution. This study is based on the assumption that the interface conformation does not necessarily correspond to the conformation shown by the solvated monomers. The primary objectives of our study were to study the interface region and analyze the hydrophobic pocket occupied by Phe87 as a potential hot spot through MD simulations. We sought to prove our hypothesis that conformational changes occurring at the dimer interface can change the active site architecture of the solvated monomers such that the monomers would no longer undergo catalysis, or if catalysis occurs, it would be at a much slower rate. We collected trajectory snapshots for further analyses from MD simulations and performed docking studies with potential small molecule disruptors. We report our results in the pages that follow.

METHODS

Protein Structure Preparation. The monomer structure used for this study corresponds to the FabH monomer (Chain B of FabH, PDB ID: 1HN9). The protein was prepared using the Protein Preparation Module available in Schrödinger's Maestro GUI.[26−29] None of the water molecules were retained; hydrogens were added, bond orders corrected, and the structure was subjected to energy minimization until the RMSD relative to the starting geometry reached 0.3 Å. This is done by a series of restrained minimizations using the IMPACT refinement module,[27−29] which alleviates potential steric clashes until the RMS deviation from the starting heavy atoms coordinates reaches 0.3 Å.

The prepared protein structure was solvated in an orthorhombic water box using the TIP4P water model. This gave a 66 Å × 74 Å × 76 Å simulation box with 11,047 solvent molecules for the monomer and a 90 Å × 90 Å × 90 Å simulation box with 21,297 solvent molecules for the dimer. The appropriate number of sodium ions was added to neutralize the net charge on the protein.

Molecular Dynamics Simulations. MD simulations were performed using the DESMOND module available from DE Shaw Research Group[30,31] using the OPLS-2005 force field. The solvated protein was subjected to energy minimization using the DESMOND minimization algorithm; either 2000 iterations or an energy gradient convergence threshold of 1 kcal/mol/Å was employed. The minimized protein was further relaxed before the actual simulation by a series of energy minimizations and short MD simulations. There are six steps in the relaxation process. The first two steps are minimization steps, solute restrained and without restraints. Steps three through six are short MD simulations of 12 ps, 12 ps, 24 ps, and 24 ps each using the NPT ensemble at 10, 10, 300, and 300 K, respectively. Velocity resampling is done in steps three through five, while step six does not include velocity resampling.

The NPT ensemble, the Nose−Hoover chain thermostat, and the Martyna−Tobias−Klein barostat were employed for the actual simulations, and the RESPA integrator was used with a time step of 0.002 ps. A 9 Å cutoff was used for the short-range Coulombic interactions, and long-range electrostatics were treated using the particle mesh Ewald method, with a

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

tolerance of $1e^{-9}$ Å. Bonds to hydrogen were constrained using the "M_SHAKE" algorithm of DESMOND. The coordinates were saved at intervals of 4.8 ps, referred to as 'frames' in this study. Coordinates of subunit B were used as the monomer for simulation, and analyses and active site analyses for the dimer refer to subunit B. The monomer and dimer simulations were carried out for comparative purposes and were sampled over 20 ns, with coordinates saved every 4.8 ps.

Distances, RMSD values, and RMSF values between residues were all calculated using the Simulation Event Analysis module of DESMOND. All frames were aligned with the starting structure prior to the calculations. The RMSD was calculated for the C_α atoms using the starting structure as reference frame, according to the formula

$$\mathrm{RMSD}_t = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(X_n - X_0)^2}$$

where X_n refers to the coordinates of the C_α atom of a particular residue at time t, and X_0 refers to the coordinates of the C_α atom of the same residue in the initial structure (Frame 0) at 0 ns, and N refers to the number of residues. Thus, RMSD at a given time t is the root-mean-square deviation of the C_α atoms of all residues at time t with respect to Frame 0. The RMSF was calculated as the root-mean-square fluctuation of C_α atoms of the individual residues over the 20 ns simulation, according to the formula

$$\mathrm{RMSF}_n = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

where X_t refers to the fluctuation of the C_α atom of a particular residue n at time t and $\bar{X}$ refers to the mean fluctuation of the residue n over the entire simulation time. Thus, RMSF of a residue is the root-mean-square fluctuation of that residue over all the time frames. The radius of gyration was calculated according to the formula

$$R_g = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(Y_n - \bar{Y})^2}$$

where Y_n refers to the distance of a particular residue from the centroid of the protein at time t, and $\bar{Y}$ refers to the mean distance of all residues from the centroid for that time t, and N refers to the number of residues.

NCI Diversity Sets. Docking studies were performed using the database of compounds available from the National Cancer Institute Developmental Therapeutics Program (http://dtp.nci.nih.gov/branches/dscb/repo_open.html). We used the NCI Diversity Sets I and II for this study. The NCI Diversity Set I is a small set of about 1900 unique compounds that recapitulates the chemical diversity of 140,000 physical compounds that are part of NCI's Plated Set chemical library. Initially, the NCI Diversity Set I was used, but in recent years, this has been superseded by the availability of the NCI Diversity Set II (and later the NCI Diversity Set III).

Ligand Preparation. The structures available from the NCI database have to be prepared for docking and enantiomers generated. The ligands were prepared using Schrödinger's Ligprep,[32] which generates tautomers, and possible ionization states at the pH range 7 ± 2 using Epik[33,34] and also generates all the stereoisomers of the compounds if necessary.

Docking. GLIDE 5.2[35] was employed for docking studies. GLIDE SP (standard precision) docking was used for initial docking runs. Further refinement of the docking poses was performed by GLIDE XP (extra precision) that incorporates additional terms in the scoring function. Ten snapshots of the monomer were selected from the 20 ns MD simulation, and these were used as protein input structures for docking to identify the most favorable conformer for binding the potential small molecule disruptors.

SiteMap Analysis. SiteMap is a tool that identifies putative binding sites by analyzing a number of parameters that contribute to tight binding of the ligand with the receptor.[36] "Site points" are identified to define a binding region and linked together depending on the closeness of the site points to the protein surface and solvent accessibility. The SiteMap program returns a number of binding site parameters: *site score*, *size*, *exposure score*, *enclosure*, *contact*, *hydrophobic/hydrophilic character*, and *donor/acceptor character*. The "size" of a binding site is approximated by the number of site points contained within a putative site. On the basis of the analysis by Halgren[37] (described later in the paragraph), the average number of site points for submicromolar sites is 132. Lower exposure scores are better, 0.52 being the average for submicromolar sites. Higher enclosure scores are good with an average of 0.76 for submicromolar sites. For contact, the average score is 1, as it is also for the hydrophobic/hydrophilic character. For the donor/acceptor character and site score, the average for the submicromolar sites is 0.76 and 1.01, respectively. *Dscore* or druggability score, provides a rough estimate of whether the site is druggable. These scores were derived by Halgren[37] by executing the SiteMap program on a number of proteins that have inhibitors bound with potencies in the submicromolar range and performing statistical analyses to produce optimized scores.[37] The OPLS-2005 force field[38] was employed, and a standard grid was used with 15 site points per reported site and cropped at 4 Å from the nearest site point.

MM/GBSA. To probe the protein−protein interactions described herein, approximate free energies were calculated using the MM/GBSA methodology. The key residue in the dimer interface, Phe87, was mutated to alanine. The binding free energies were calculated using MM/GBSA for both the mutated and wild type protein, and the results were compared. MM/GBSA uses an all-atom method, which uses molecular mechanics and continuum solvent and the OPLS_2005 force field to calculate the binding free energies. In this approach, the free energy is calculated according to the following equation

$$\Delta G_{\mathrm{binding}} = G_{\mathrm{complex}} - (G_{\mathrm{receptor}} + G_{\mathrm{ligand}})$$

MM/GBSA calculates binding free energies similar to MM/PBSA (Molecular Mechanics Poisson−Boltzmann Surface Area) that uses the Poisson−Boltzmann model to calculate the electrostatic solvation energy as opposed to MM/GBSA, which uses the generalized Born model to calculate the electrostatic solvation energy. The accuracy and performance of MM/GBSA vs MM/PBSA were recently reported by Wang et al.,[39] which revealed that in most systems MM/GBSA performed better than MM/PBSA in predicting relative binding free energies.[39]

RESULTS AND DISCUSSION

As previously stated, the dimerization interface of FabH in *E.coli* is approximately 3200 Å^2 as revealed by its crystal

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

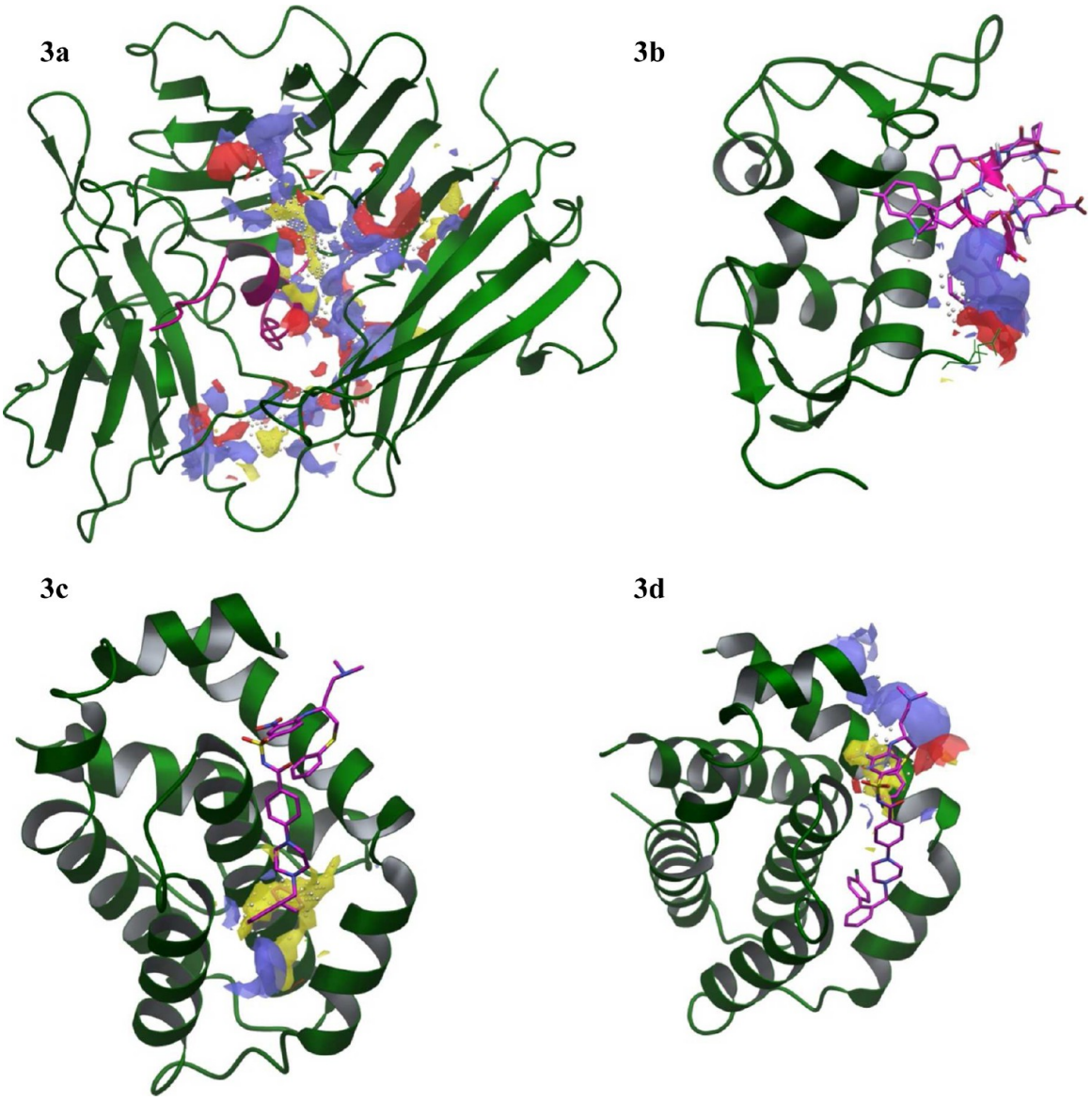


Figure 3. (a) Binding site of the IL-1 receptor[44] (PDB ID: 1G0Y) analyzed by SiteMap; the native peptide is colored magenta. (b) Binding site of MDM2[46] (PDB ID: 2AXI) analyzed by SiteMap; the native peptide is colored magenta. (c,d) Top two binding sites identified by SiteMap on the protein Bcl-xL[48] (PDB ID: 2XYJ); the native ligand (ABT737) is colored magenta; The binding sites are represented as hydrophobic, H-bond donor, and H-bond acceptor maps (blue, hydrogen-bond donor; red, hydrogen bond acceptor; yellow, hydrophobic).

structure (PDB ID: 1HN9[17]). This structure was chosen because of its high resolution compared to the earlier ones. This structure does not have any ligand bound (except for phosphate, which is distal to the active site and does not appear to play a catalytic role), which is important to understand the dynamics of the unliganded system. However, this orthogonal space group structure is quite similar to the reported structures with ligands bound.[18,19] The key residues in the dimer interface are oriented toward the surface of the monomer in the interface region.[40] To probe the dimer interface as a probable binding region, SiteMap[41] analysis was performed on the monomer and dimer. We assumed that this analysis might also help identify the hot spots present in the dimer interface region.[42] Because the dimer interface of FabH is a *bona fide* protein−protein interaction (PPI), we validated the utility of SiteMap as a drug discovery tool for PPIs by performing SiteMap analysis on proteins that are known to be involved in PPIs and for which there are X-ray structures with relatively small molecules bound to the PPI interface of one of the protein partners. In the current investigation, the proteins chosen were the Interleukin-1 receptor (IL-1R), which interacts with IL-1β,[43,44] MDM2, which interacts with the transactivation domain of p53,[45,46] and Bcl-xL, which interacts with the BH3 domain of proteins such as Bax and Bak.[47,48] X-ray structures of all of these proteins have either a peptide or a non-peptidic ligand bound to the interface region to which their protein binding partner binds (Figure 3).[22] An analysis was performed to determine whether SiteMap could identify the binding sites occupied by the native small molecule or peptide bound to the respective proteins which, of course, represents the protein−protein interaction (PPI) interface. In all cases, SiteMap identified the binding site within the top two site rankings (Table 1).

The binding regions of Inter-Leukin-1 receptor[44] and MDM2[46] had site scores of 1.05 and 0.92, respectively. For Bcl-xL, the binding region of **ABT737** (a known inhibitor for Bcl-xL) was identified as two different sites with site scores of

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. SiteMap Analysis on FabH (Monomer and Dimer) and Other Proteins Involved in PPIs

title	SiteScore[c]	size[d]	*Dscore*[e]	volume[d]	exposure[d]	enclosure[d]	contact[g]	hydrophobic[f]	hydrophilic[f]	don/acc[I]
Interleukin Receptor-1	1.05	427	1.03	1100.34	0.43	0.77	1	0.59	1.15	1
MDM2	0.92	39	0.98	99.13	0.59	0.73	0.92	3.32	0.13	5.87
Bcl-xL-Site1	0.94	48	0.99	128.63	0.56	0.74	0.93	2.78	0.34	11
Bcl-xL-Site2	0.71	43	0.73	114.91	0.75	0.51	0.65	0.77	0.6	2.75
Monomer-Site1[a]	1.1	83	1	112.5	0.35	0.94	1.3	1.06	1.31	0.13
Monomer-Site2[b,a]	1.01	103	1.07	288.81	0.69	0.64	0.81	0.84	0.77	0.98
Dimer-Site1	1.1	82	1	112.16	0.35	0.95	1.29	1.04	1.28	0.13
Dimer-Site2	1.04	104	1.04	210.95	0.48	0.76	1.01	0.79	1.07	0.24
Dimer-Site3[b]	1.07	267	1.02	485.69	0.49	0.8	1.04	0.58	1.24	0.83
Dimer-Site4[b]	0.86	63	0.91	204.43	0.77	0.59	0.68	0.71	0.58	1.11
Dimer-Site5[b,a]	0.84	72	0.84	251.76	0.75	0.62	0.71	0.25	1.01	1.12

[a]Used in this study. [b]Dimer interface binding sites identified by SiteMap. [c]Used to identify and compare binding sites; a score of 0.8 or higher is generally considered a binding site. [d]Number of site points for the given binding site; a size of 132 or higher is optimal for submicromolar sites. [e]Criterion to decide whether or not the site is druggable; a score greater than 0.98 is considered "druggable". [f]Refers to the openness of the binding site to the solvent; lower exposure scores and higher enclosure scores are considered better. [g]A measure of vdW non-bonded interactions of the site points with the receptor; average score for submicromolar sites being 1.0. [h]Refers to the hydrophobic and hydrophilic properties of the binding site, respectively; average scores are 1.0 for submicromolar sites. [I]Refers to the ability of the ligand to donate or accept hydrogen bonds; average score for submicromolar sites is 0.76.

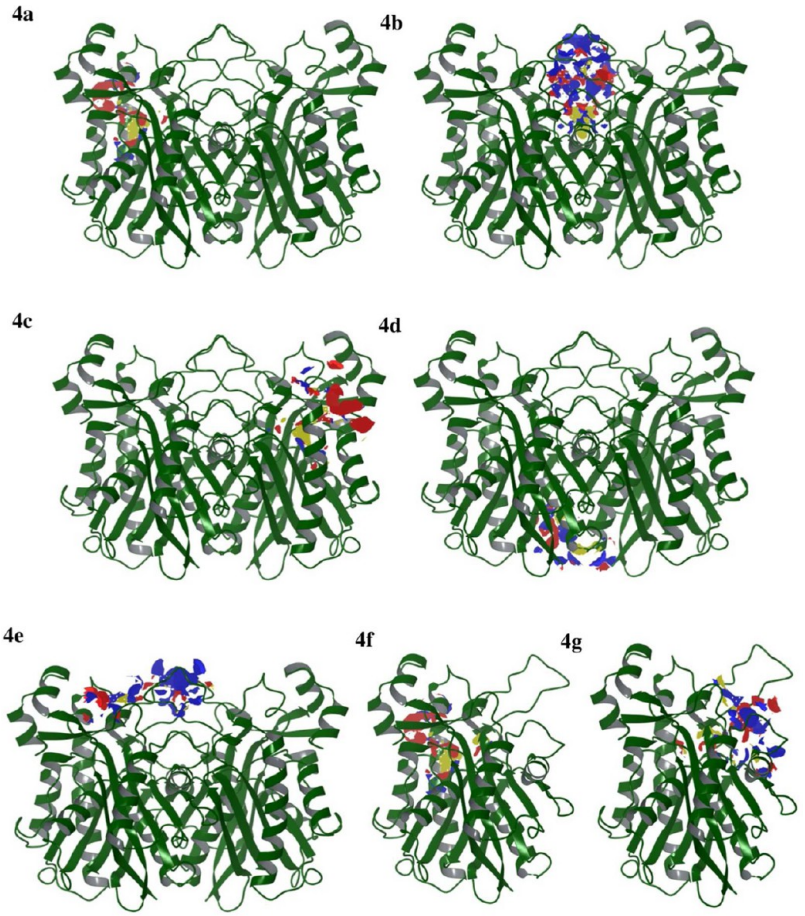


Figure 4. Potential binding sites as identified by SiteMap on the dimer and monomer FabH. Analysis was performed on the crystal structure (PDB ID: 1HN9). Protein represented as a cartoon; binding sites are represented as a donor−acceptor hydrophobic map. (a) Dimer-site1, (b) Dimer-site3, (c) Dimer-site2, (d) Dimer-site4, (e) Dimer-site5, (f) Monomer-site1, and (g) Monomer-site2.

0.94 and 0.71, respectively. Bcl-xL has four hydrophobic pockets p1-p4 that are occupied by the substituents of the molecule **ABT737**.[48] Hydophobic pockets p1-p2 are located close to each other, and p3−p4 are located close to each other with solvent accessible region between them. Each pairing of pockets was found as a single binding region in the SiteMap analysis. The site scores were well within the ranges described by Halgren.[37]

SiteMap analysis performed on monomeric and dimeric FabH is also tabulated in Table 1. SiteMap identified five different binding sites for the FabH dimer. For dimeric FabH, as depicted in Table 1 the two biological active sites, Dimer-

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Site1 and Dimer-Site2, were identified within the individual monomers. Dimer-Site3 was found to have the largest area in the dimer interface region, while Dimer-Site4 and Dimer-Site5 occupied shallow binding pockets in the dimer interface region. For monomeric FabH, SiteMap identified two probable binding sites; Monomer-Site1, which is the biological active site, whereas Monomer-Site2, which spans residues 189−193, is the dimer interface and is used in this study (Figure 4). The dimer interface region was identified as a potential binding site in both the monomer and the dimer with a site score greater than 1. A *Dscore* of 1.02 for the dimer interface binding region in the FabH dimer implies that the site is "druggable" according to Halgren et al.[37] (where values greater or equal to 1 are favored). This binding region has a relatively low exposure score and relatively high enclosure score, which is optimal for a submicromolar binding site. The donor/acceptor character of the dimer interface binding sites in both the monomer and dimer are also close to the optimal values. These findings strongly suggest that the Monomer-Site2 defines a hot spot,[42] which is further substantiated by the fact that it has been reported that the residues which comprise this interface region (Phe87 and its complementary residues from the adjacent monomer; Thr190, Leu191, Pro192, Asn193) are important for the stability of the FabH dimer.[49]

Findings from the SiteMap analyses reveal that the dimer interface in FabH could be an interesting allosteric site of inhibition. This fact could be exploited in drug design by creating an inhibitor that could lock FabH into its monomeric state and hence render the protein significantly less active than when in the dimeric state as discussed earlier. In order to design inhibitors of this type, detailed knowledge of the dimer interface structure for the monomer in solution is essential. To probe the conformational space accessible to this dimer interface in solution, a 20 ns MD simulation of the FabH monomer was carried out. For comparison purposes, we carried out a 20 ns MD simulation on the FabH dimer.

Molecular Dynamics. Protein structures must be converged in order to achieve predictable results from an MD simulation, and root-mean-square deviation (RMSD) is a good indicator of overall stability of any protein system. RMSD of the C_α atoms were followed over the 20 ns time period (Figure 5). All the analyses were performed using the Simulation Event

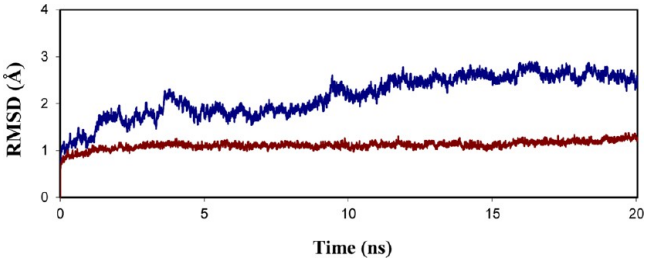


Figure 5. RMSD for the monomer (blue) vs Dimer (red) in *E. coli* FabH over the 20 ns MD simulation.

Analysis (SEA) tool available in DESMOND.[30,31] In the dimer, the protein was found to be fairly stable during the entire simulation with a mean RMSD of 1.12 ± 0.08 Å, while the monomer converged after 10−12 ns and had a mean RMSD of 2.52 ± 0.44 Å (Figure 5). Radius of gyration is another indicator of protein stability. The radius of gyration for the monomer and dimer during the 20 ns simulation time was followed, which indicated both the systems to be stable (Figure 6).

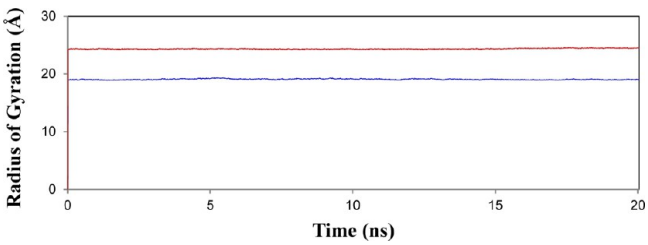


Figure 6. Radius of gyration measured for the monomer (blue) and the dimer (red) in *E. coli* FabH over the 20 ns MD simulation.

Root mean square fluctuation (RMSF) is a measure of the average fluctuation of the residues over time. RMSF of the C_α atoms of all the residues over different time frames were plotted to determine the convergence of individual residues and analyze the fluctuations of dimer interface residues and active site residues over the simulation time. The monomer was found to have converged after 10−12 ns (Figure 7a), while the fluctuations were relatively minimal for most residues throughout the dimer simulation (Figure 7b). These results correlate well with the RMSD data (Figure 5). For comparison purposes, in Figure 7c the RMSFs of the monomer and dimer over 20 ns are depicted at the same scale. The residues that had large fluctuations were the ones in the dimer interface region or residues adjacent to the active site. The fluctuations of residues in the dimer interface region in the monomer were almost twice as much as the dimer. These results are consistent with large conformational changes undergone by the solvated monomers compared to the dimer.

An RMSF difference plot was also generated by plotting the difference in RMSF of the monomer compared to the dimer. This was done to determine the regions that had significant structural changes in monomer vs dimer during the simulation. Both subunits of the dimer were used for this plot (Figure 7d). The RMSF difference plot using Chain A and the monomer is very similar to the RMSF difference plot using Chain B and the monomer. In the RMSF difference plot (Figure 7d), any difference in fluctuation above the ±0.5 Å region was considered significant.

Residues that had large fluctuations were residues 31−38, 86−89, 129, 148−153, 191−208, 224, and 253−265. These regions are labeled 1−6 in the RMSF plots, and the regions corresponding to those residues in the plot are shown in Figure 7e, except residue 224. Residue 224 was not included in the discussion because it was the only residue that displayed a negative RMSF difference; however, its flexibility could be explained because of its proximity to the active site. Peak 1 consists of residues 31−38 that lie in the entrance to the biological active site and shows large motions. The RMSF difference plot for this peak is smaller for Chain B by about 0.5 Å when compared to Chain A in the dimer, which might be attributed to the asymmetric movement of the subunits in the dimer (Figure 7d). Peak 2 contains the Phe87 residue that is solvent accessible in the monomer but is buried in a hydrophobic pocket in the dimer and hence shows large fluctuations in the monomer while being relatively static in the dimer. This peak is not present in the dimer RMSF plots. Peak 3 encompasses residues 126−131, which are present at the dimer interface. Peak 4 involves residues 148−153, which lie at

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

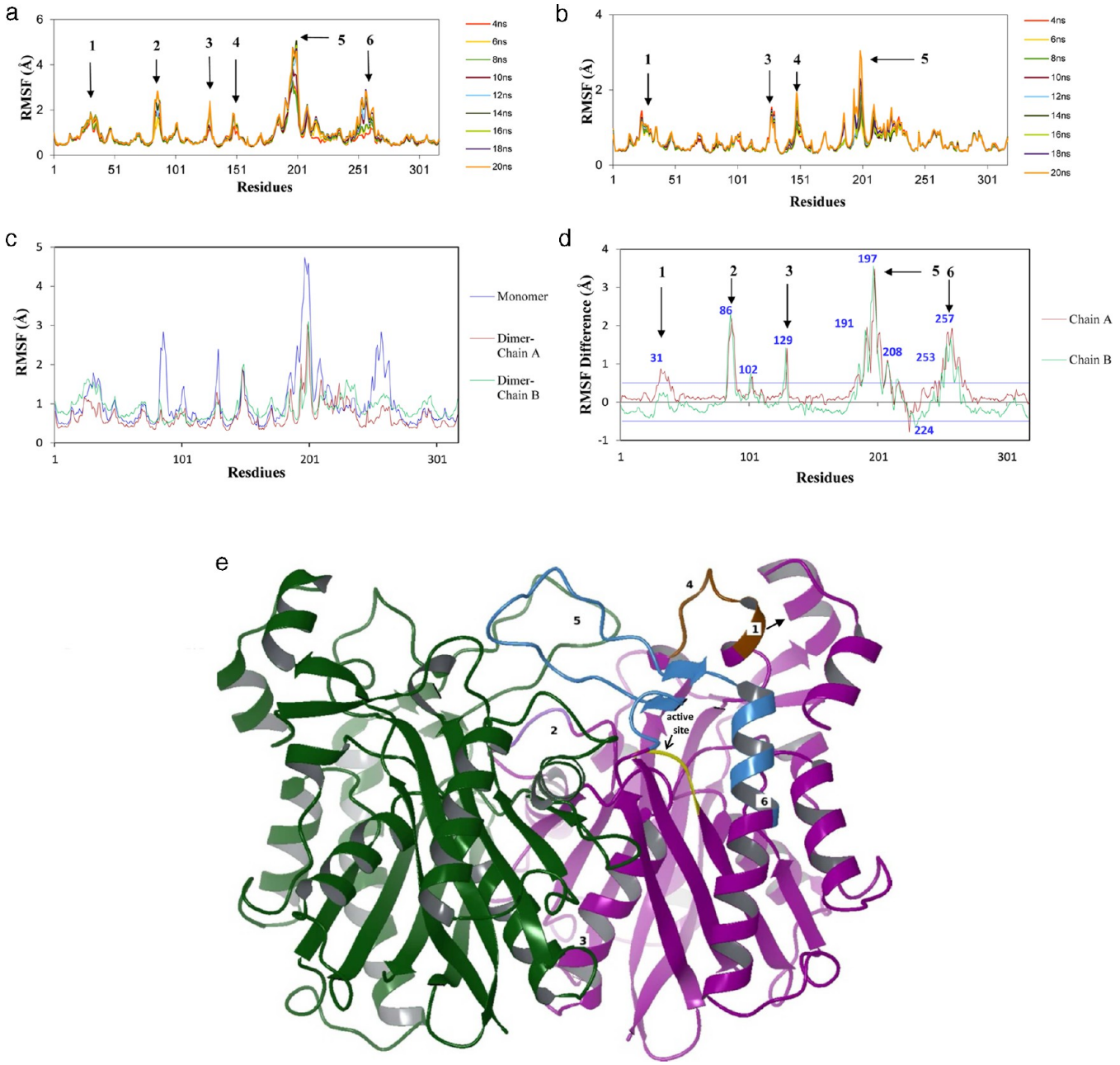


Figure 7. RMSF plots of monomeric (subunit B) and dimeric FabH. (a) RMSF over different time frames for FabH monomer. (b) RMSF over different time frames for FabH dimer (subunit B). (c) RMSF of monomer and dimer averaged over the entire 20 ns simulation (red, dimer, blue, monomer; the catalytic cysteine is labeled). (d) RMSF difference plot of monomer vs dimer over the 20 ns simulation. Note: peak 1 consists of residues 31−38, which lie in the entrance to the biological active site. Peak 2 contains residue Phe87. Peak 3 encompasses residues 126 through 131, which are present at the dimer interface. Peak 4 involves residues 148−153, which lie at the entrance to the biological active site. Peak 5 consists of residues 181−217, which form part of an α-helix and a large loop that extends over the dimer interface region. Peak 6 consists of residues 253−265, which are close to the catalytic triad. (e) FabH dimer and the peak labels corresponding to the RMSF plots.

the entrance to the biological active site and exhibit large fluctuations. Peak 5 consists of residues 181−217, which form part of an α-helix and a large loop that extends over the dimer interface region. Peak 6, which is also absent in the dimer, consists of residues 253−265 that are present closer to the catalytic triad and show large fluctuations in the monomer thereby changing the active site architecture, while being stable in the dimer. These results confirm our hypothesis that conformational changes occurring in the solvated monomer due to lack of association with a second monomer through the dimer interface, ultimately result in large active site architectural changes making it unlikely that the solvated monomers would be catalytically competent.

Principal Component Analysis (PCA). PCA is an efficient analysis tool to reduce the number of variables and analyze conformational changes along the principal axes of motion. The major motions of the protein system are generally captured in the first few principal components. The covariance matrix was generated from the DESMOND trajectory based on the coordinates of the C_α atoms of each residue. The covariance matrix using every 10th step of the trajectory for the monomer and dimer systems was generated using the trajectory_covar-

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

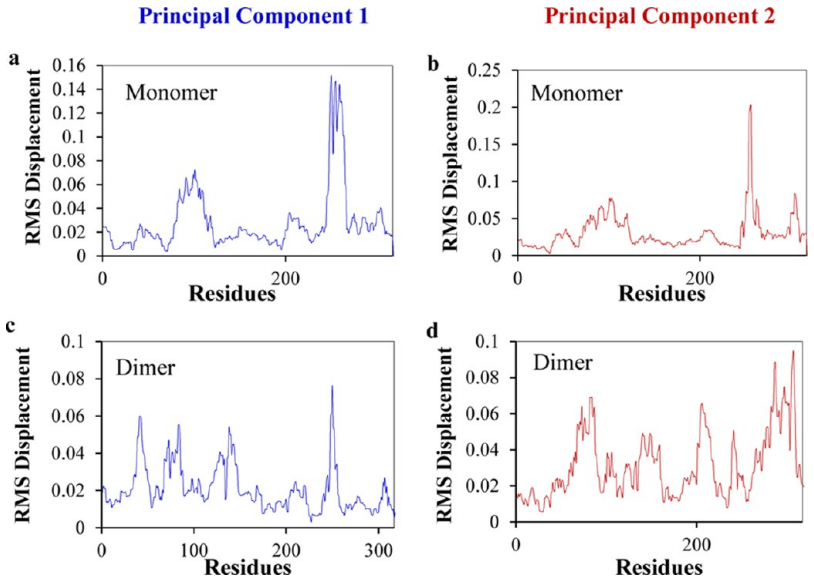


Figure 8. PCA plots. (a) PC1 plot of monomer. (b) PC2 plot of monomer. (c). PC1 plot of dimer. (d) PC2 plot of dimer. The PCA methodology used is described in the Supporting Information.

iance_matrix.py script of DESMOND, which generated a 1902*1902 (3N*3N) matrix for the dimer and a 951*951 (3N*3N) matrix for the monomer. PCA was performed on the covariance matrix using the MATLAB 2010 environment, and the graphs were plotted in Excel.

The plots of the first two principal components (PC1 and PC2) of the monomer and dimer are shown in Figure 8, while plots of principal components 3−5 (PC3, PC4, and PC5) are shown in the Supporting Information. The first 5 PCs accounted for 78% of the motion in the monomer and 58% in the dimer. As observed in the RMS displacement plot for the monomer (Figure 8a,b), PC1 and PC2 both show fluctuations in the regions containing residues 78−111 (overlapping with peak 2 in RMS displacement plot of monomer; Figure 7a) and 244−267 (overlapping with peak 6 in RMS displacement plot of monomer; Figure 7b). In the dimer, PC1 shows fluctuations in the regions containing residues 32−50, 66−89, 117−127, 134−148, and 240−246 (Figure 8c). PC2 shows fluctuations in the regions containing residues 59−90, 126−160, 190−219, 235−245, and 271−310 (Figure 8d). Comparing the scale of monomer vs the dimer PC plots, fluctuations in the dimer are minimal and are approximately the same value for all the peaks, indicating that the fluctuations are not as significant as in the monomer. These plots correlate well with the RMSF plots shown in Figure 7 and again confirm our hypothesis that the changes occurring in dimer interface of the solvated monomer are correlated with conformational changes at or near the biological active site.

Biological Active Site. In FabH, the active site is composed of a catalytic triad formed by residues, Cys112, His244, and Asn274 (Figure 9a).[18] Residue Cys112 is involved in acetyl transfer, while His244 and Asn274 play an active role in decarboxylation.[17] Residues His244 and Asn274 are present close to Cys112 in the long active site tunnel, such that the distance between the $C_{\alpha H244}-C_{\alpha N274}$ atoms is ∼8 Å. Cysteine 112 lies on the bottom of the active-site tunnel, the entrance of which is flanked by residues Trp32 and Arg151, as discussed earlier.[6] The per-residue RMSD plot over the 20 ns simulation time shows that the catalytic residue Cys112 fluctuates with a maximum RMSD of 24.6 ± 4.8 Å in the monomer (Figure 9b). In the dimer, this catalytic residue shows greater fluctuations with a maximum observed RMSD of 35.2 ± 7.5 Å for the 20 ns simulation. This result suggests that the catalytic residue fluctuates in order to accommodate the binding of the substrate, and it would be prudent to include induced fit effects when docking inhibitors in the active site.[50] There are numerous examples of catalytic residue fluctuations to accommodate substrate binding. For example, the penicillopepsin protein (PDB ID: 1BXO) has residue Tyr75 at the tip of a β-hairpin loop that forms the catalytic pocket, which shows large fluctuations to accommodate the docking of inhibitors.[50] Interestingly, our analysis shows that the fluctuations for Cys112 are minimal in the monomer as compared to the dimer. In the monomer, the active-site conformation is changed and presumably can no longer accommodate Acetyl-CoA or substrate.

Distance Plots (Catalytic Triad). The distances between the C_α atoms of His244, Cys112, and Asn274 were calculated during the 20 ns simulation to understand the dynamics of the catalytic residues. In the dimer, the distance between Cys112 and the other two catalytic residues remains between 9 and 10 Å. In the monomer, a significant difference could be observed for both the distances. The distance between Cys112 and His244 remains around 9 Å for the first 3.5 ns but is drastically lowered by ∼1.5 Å for the rest of the simulation (Figure 9c). The distance between Cys112 and Asn274 stays at ∼9 Å for the first 3.5 ns but increases quickly by ∼1.5 Å for the remaining simulation time (Figure 9d). This shows that the biological active site has undergone a very significant conformational change; Cys112 and His244 came closer to each other, while Cys112 and Asn274 drifted away changing the architecture of the tunnel-like active site.

The mechanism of FabH proposed by Qiu et al.[18] suggests a proton transfer between the thiolate hydrogen of Cys112 and the N_α nitrogen of His244, which generates the thiolate anion that attacks the carbonyl carbon of acetyl CoA (Figure 9e) and a final proton transfer between N_α nitrogen of H244 that is protonated and a thiolate anion of Cys112 after condensation. If such was the case, then the two atoms involved in proton transfer would be expected to be at an optimum distance to aid

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

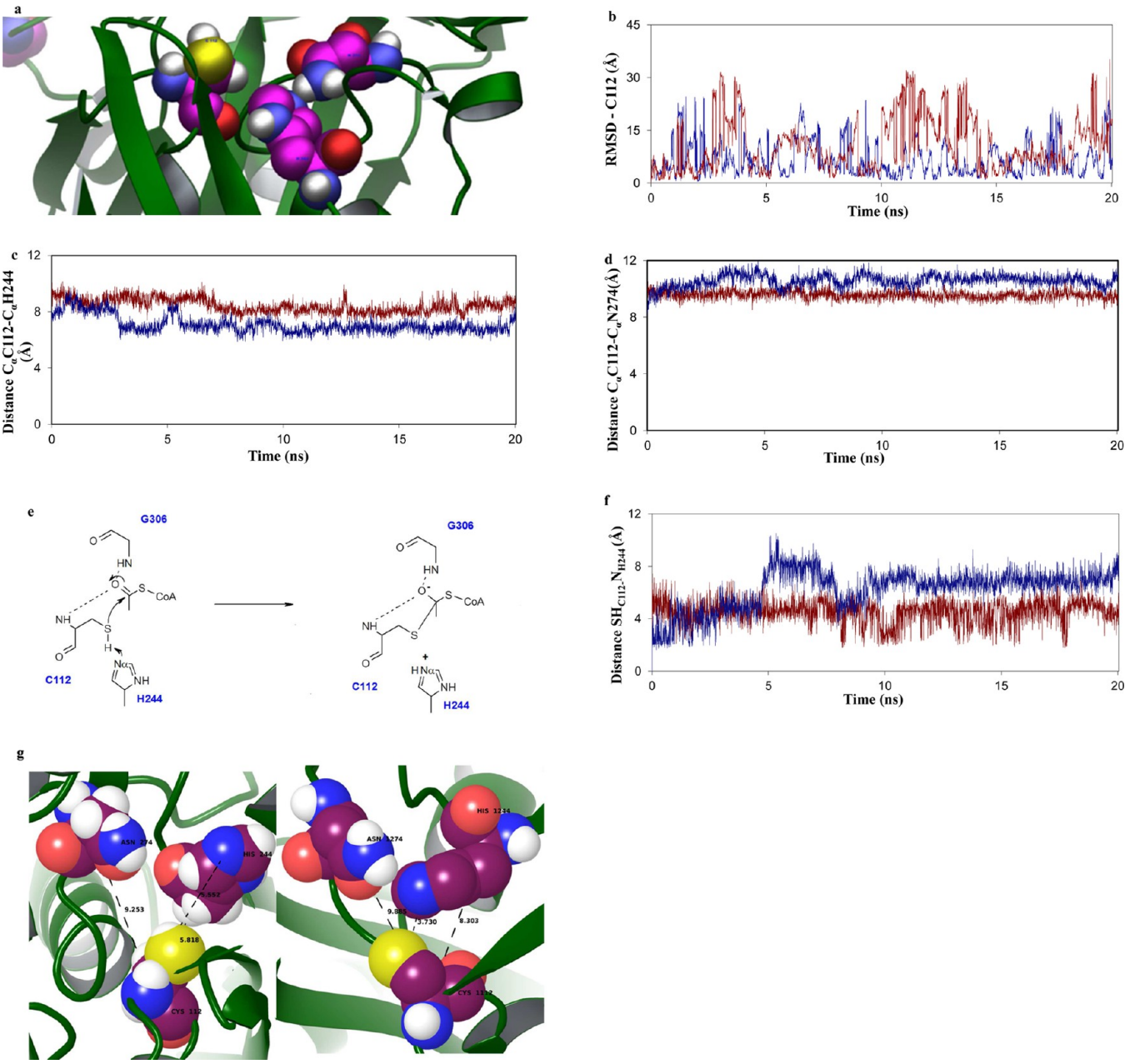


Figure 9. Distance plots for the catalytic triad. (a) Active site tunnel of FabH (represented as ribbon) showing the catalytic triad, Cys112, His244, and Asn274 represented CPK and colored coded according to elements present. (b) RMSD plot for all atoms in residue Cys112 for 20 ns simulation. (c) Distance plot for the C_α atoms between Cys112 and His244. (d) Distance plot for the C_α atoms between Cys112 and Asn274. (9) Proposed mechanism of proton transfer between His244 and Cys112 as proposed by Qiu et.al.[17] (f) Distance plot between Cys112-SH proton and sp2-hybridized nitrogen of His244. (g) Active site tunnel of FabH in the average trajectory structure of the monomer (left) and dimer (right) showing the average distances between the C_α atoms of Cys112 and His244 and Cys112 and Asn274 and the average distance between the thiolate hydrogen of Cys112 and N_α nitrogen of His244 during the 20 ns simulation.

the transfer. The distance between these atoms in the crystal structure is 3.7 Å. We followed the distance between the thiolate hydrogen in Cys112 (-SH) and the N_α nitrogen in His244 (N) (Figure 9f) to determine if it changes in the solvated monomers because the active site has undergone large conformational changes. This distance starts around 3 Å in the monomer and remains around that value for the first 3.5 ns before increasing by a stark instantaneous value of 5 Å and a value of approximately 6 Å for the rest, while the distance remains around 4 Å for the dimer. A distance of 4 Å could be argued to be a reasonable distance between the two atoms involved in proton transfer as is the case with the dimer. In the monomer, this value shifts to around 6 Å indicating that the side chains of these two residues swing away from each other due to a conformational change of the active site. In fact, the solvated monomers undergo a huge conformational change in the active site caused by the increased proximity of the C_α atoms of Cys112 and His244, which in turn moves the side chains of these two residues away from each other, thereby increasing the distance between the atoms involved in proton transfer. This suggests that in the monomer, if catalysis does occur, would do so at a much slower rate than the dimer.

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Distance Plot of Trp32 and Arg151. The biological active site of FabH occupies a long cylindrical tunnel as explained earlier. The side chains of residues Trp32 and Arg151 are reported to open and close upon binding the substrate.[6] The distance between these two residues is an important factor to be considered because it determines the ligand binding affinity within the biological active site. The distance between the C_α atoms of these key residues was investigated over the course of the simulation in the monomer and dimer. The distance between the C_α atoms of these two residues in the monomer is found to fluctuate by a greater degree than the dimer (Figure 10). Moreover, the average distance in the monomer is 2.3 Å

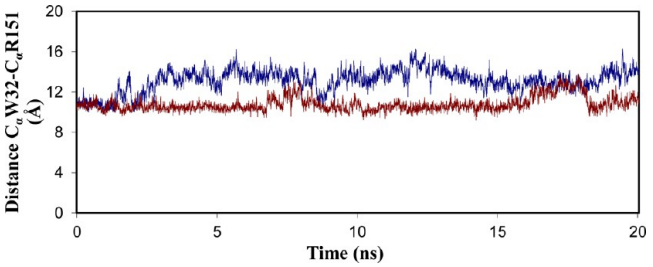


Figure 10. Distance between the residues Trp32 and Arg151 in the monomer (blue) and the dimer (red) in *E. coli* FabH over the 20 ns MD simulation.

greater than for the dimer, which suggests tighter substrate binding in dimer. The initial distance between the two residues is around 2 Å in the closed form, which could be a tight binding conformation. In the open form, the distance between the residues increases, making the active site more accessible but not necessarily tight binding.

Torsion Angles. In the next step, we wanted to analyze the dynamics at the dimer interface, especially the hydrophobic pocket occupied by Phe87, as a potential hot spot in the dimer interface region (Figure 11). Large fluctuations in torsion angles are, of course, indicative of large conformational changes. Residues (Ala83, Thr84, Ala86, Phe87, Asp107, Ala109, Tyr 125, Asp182, Ser184, Thr190, Leu191, Pro192, Asn193, and Ile203) were analyzed for torsion-angle fluctuations during the 20 ns simulation in the monomer and dimer. Two different plots were recorded for the dimer, one corresponding to each subunit. Plots for Chain B are shown here; however, plots for subunit A (Chain A) are similar to plots of Chain B, with the exception of residues Ala83, which show less fluctuations in Chain A and Asn193, which shows more fluctuations in Chain A (shown in SI). The phi and psi angles were measured using simulation event analysis of DESMOND, and the dials plots (Figure 12) were recorded using the "R" language and environment for statistical computing[51,52] with an in-house script.[53] In the monomer, residues Ala83, Thr84, Ala86, Ala109, Asp182, Ser184, Thr190, Leu191, and Asn193 show large fluctuations from the initial value. In Figure 13, Ramachandran plots were recorded as scatter plots in Excel (initial position shown as a green dot in the scatter plot). In the monomer, the side chains of these residues appear to be highly flexible during the MD simulation, which has a dramatic effect on the hydrophobic pocket in which residue Phe87 is embedded. Residues Asp107 and Tyr125 that lie in the ß-sheet and left-handed α-helix regions show localized fluctuations compared to the residues in a loop (Figure 12).

The deviation for the two angles of the loop residues from the initial value indicates that the starting conformations of these residues are not representative for the monomeric state in aqueous solution (Figure 13). Residues Tyr125, Asp182, Asn193, and Ile203 greatly deviate from the initial orientation, the dihedral angles shifting almost 60−120° degrees from the initial position.

Trajectory Clustering. To study the conformational changes in the dimer interface region of the monomer, trajectory clustering on the monomer was performed using DESMOND.[30,31] Trajectory clustering was performed using the "single" linkage method based on the backbone RMSD in DESMOND. The analysis identified the top 10 clusters. The representative members being Frames 160, 580, 820, 1260, 1700, 2020, 2330, 2640, 3420, 4140 corresponding to simulations times of 0.77, 2.78, 3.94, 6.05, 8.16, 9.70, 11.18, 13.25, 16.42, and 19.87 ns, respectively. Conformational changes occurring in the biological active site are shown in Figure 14. Each of the snapshots is superimposed with Frame 0 for easy visualization of conformational changes in the residues shown. Conformational changes occurring in residues comple-

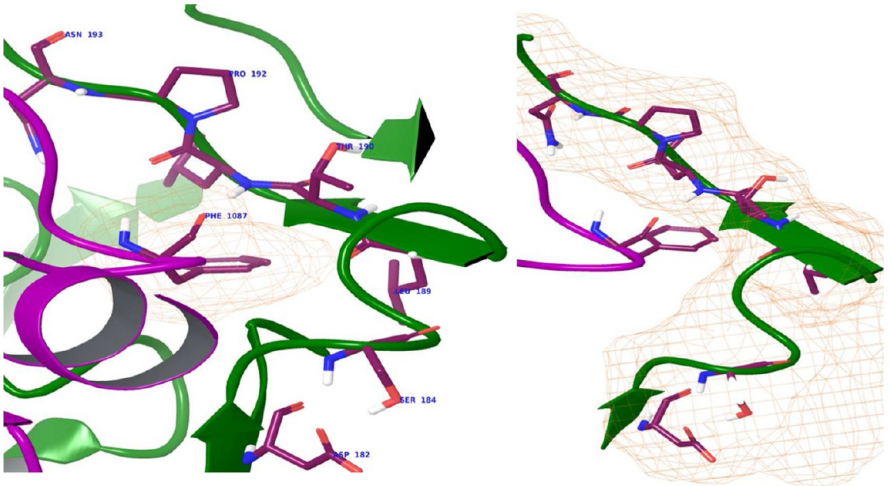


Figure 11. Hydrophobic pocket formed by residue Phe87 and complementary residues from the other monomer. Molecular surfaces generated for Phe1087 (left) and complementary residues (right) shown separately in the two images for clarity.

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

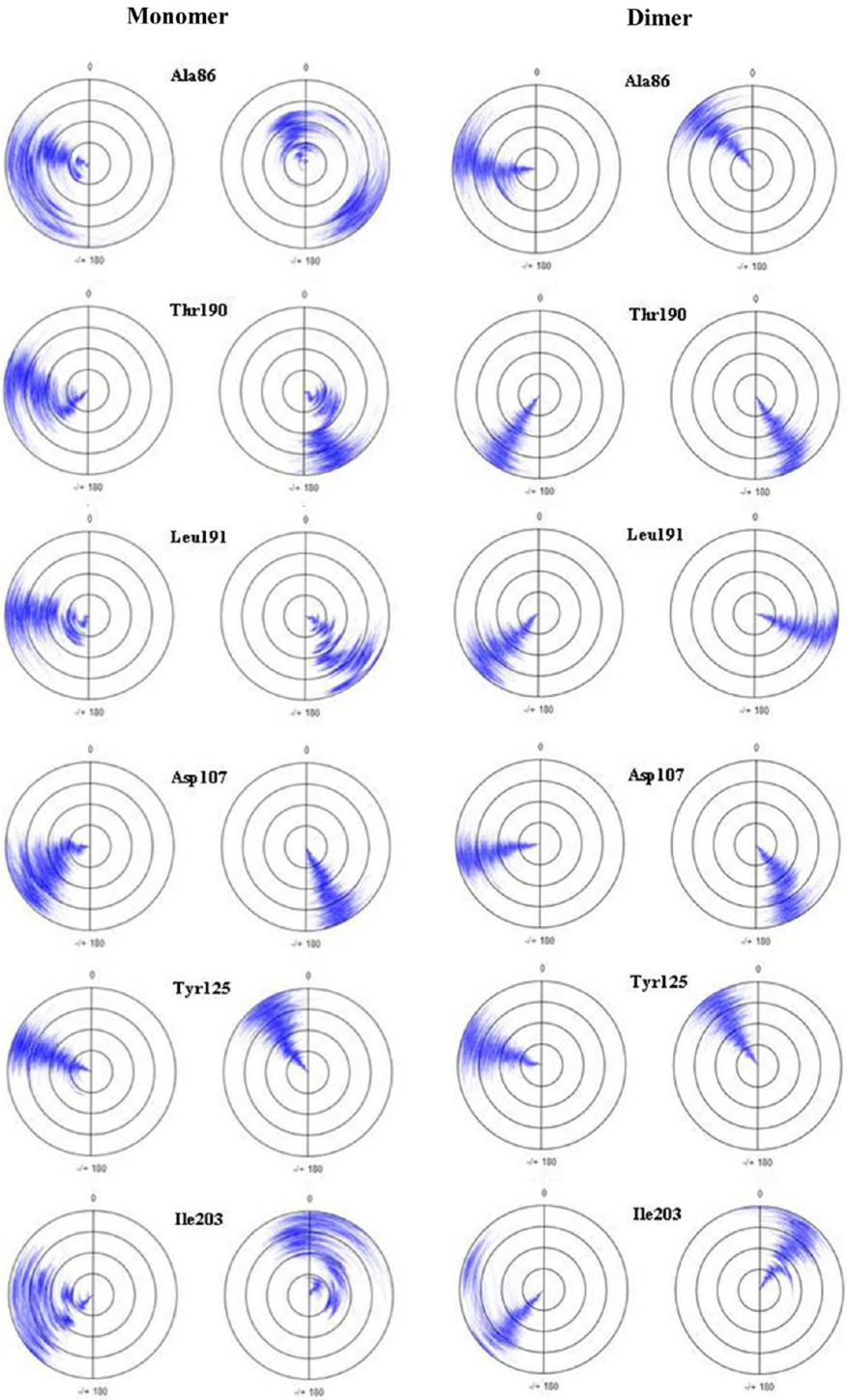


Figure 12. Representative dialsplots of the ϕ and ψ angles of the residues in the interface region for the monomer in *E. coli* FabH over the 20 ns MD simulation. The center point corresponds to 0 ns, and the edge corresponds to 20 ns. Each residue has two plots; first plot represents the ϕ angle, while the second plot represents ψ angle.

mentary to the Phe87 pocket are illustrated in the Supporting Information.

SiteMap analysis was performed on the representative members of each cluster to identify the binding regions in each. They were compared to the SiteMap results from Frame 0 (Frame 0 corresponds to the starting structure of the monomer before MD was performed).

A summary of the SiteMap analyses is shown in Table 2 for Monomer-Site2. SiteMap analyses on the representatives of the cluster identified the dimer interface (Monomer-Site 2) as the major binding region as opposed to the active-site in most cases.

Alanine Scanning and MM/GBSA Studies. To further validate the hydrophobic region involving Phe87 as a hot spot in the dimer interface region, alanine mutagenesis studies were performed on the Phe87 residue. Mutation of Phe87 to alanine should result in an inactive enzyme because Phe87 appears to be means of significant "communication" between the

dx.doi.org/10.1021/ci3003528 | *J. Chem. Inf. Model.* XXXX, XXX, XXX–XXX

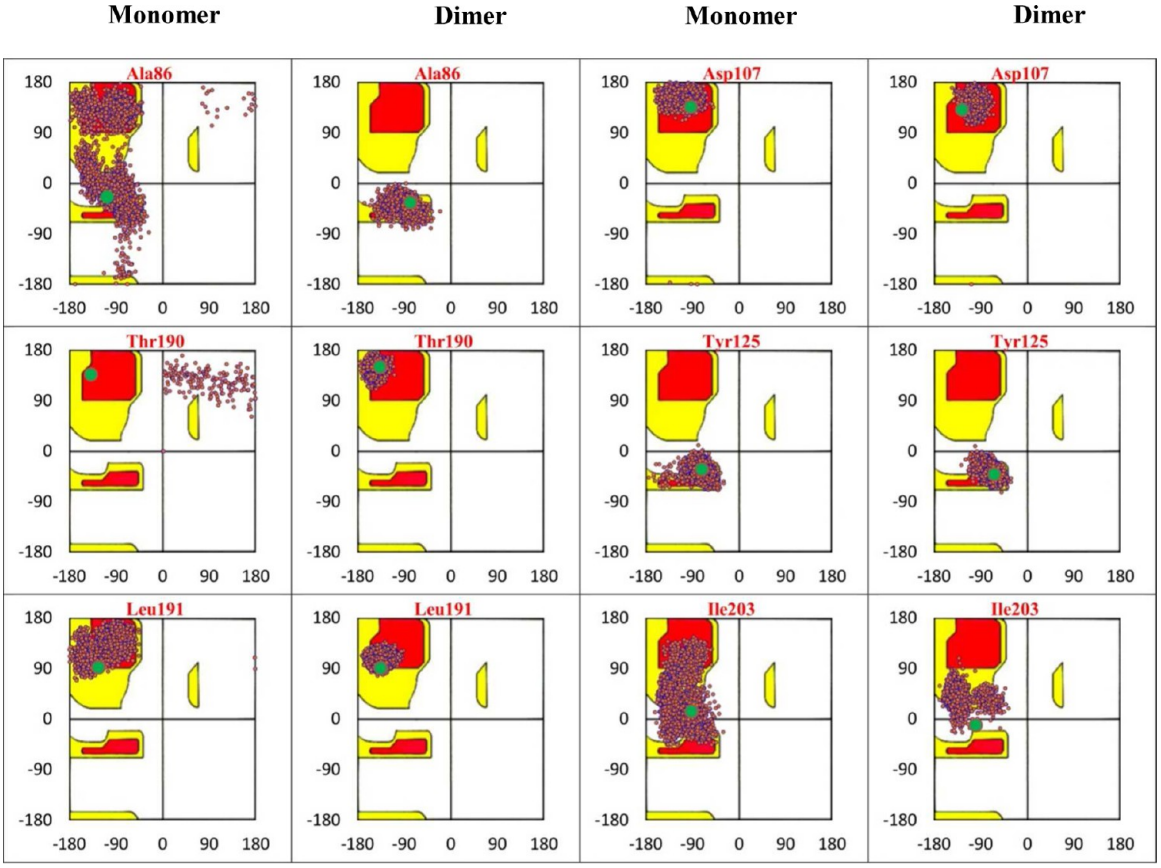


Figure 13. Scatter plots of the ϕ and ψ angles of representative residues in the interface region for the monomer in *E. coli* FabH over the 20 ns MD simulation. The starting structure dihedrals are represented by a green dot. Angles are plotted on a Ramachandran plot background to clearly identify the β-sheet and alpha helical regions.

individual monomers.[18] To analyze the importance of this residue at the dimer interface, we mutated one of the Phe87 residues to alanine (F87A in one monomer and F87 in the other monomer) and performed a 6 ns MD simulation to compare with wild type (WT). Thus, MD simulations were run on the mutant (entire dimer with F87A instead of F87 in one of the monomers) and on the wild type protein for 6 ns. Trajectory snapshots were collected every 0.5 ps for both the WT and the mutant. MM/GBSA studies require a docked conformation of the receptor and ligand.[35] Because docking the entire monomer (317 residues) was feasible with GLIDE due to a 300 atom restriction for the ligand, the monomer was truncated to a 7-mer and was treated as the ligand for rigid docking. For WT, the monomer containing the Phe87 (which was mutated to alanine in the mutant) was truncated to a 7-mer peptide centered on Phe87. Similarly, for the mutant, the monomer containing F87A was truncated to a seven residue peptide: residues THAFPSA for wild type and THAAPSA for mutated protein (Figure 15).

The truncated peptide was then prepared using the protein preparation module of Schrödinger. GLIDE refinement was performed to get the preparatory files for MM/GBSA studies, which requires the protein and peptide to be in a reasonable docked conformation. Unlike other docking options, the "refine" option of GLIDE does not dock the ligand but just refines the ligand in the binding site by performing energy minimization of the ligand structure in the field of the receptor. Refinement was performed using the standard precision (SP) method in GLIDE.[35] For GLIDE refinement, the truncated 7-mer peptide was considered as the ligand, and the other monomer was considered as the receptor and the dimer interface region as the ligand binding region. MM/GBSA calculations were then performed on each of the GLIDE output files. The approximate binding free energies (ΔG) obtained from the MM/GBSA[39,54] studies are listed in Table 3 and represent ΔG values for the binding of the wild type and mutated 7-mer peptides derived from the second monomer.

A difference in ΔG between the wild type (WT) and mutant were calculated and averaged over the 13 selected snapshots. It can be inferred that the alanine mutation was disfavored or the WT was favored over the alanine mutant by −21.56 kcal/mol as expected. This finding is extremely important in our study because the hydrophobic pocket into which residue Phe87 is embedded forms a potential inhibitor binding pocket at the dimer interface; therefore, this area is a potential hot spot.

Virtual screening provides an interesting way to compare the approximate binding free energies of different ligands in the field of the receptor. Rigid docking to the FabH interface binding pocket defined by Phe87 was performed using GLIDE SP. As proof of concept and to further validate the dimer interface as a potential target, we performed comparative docking studies with the MDM2 (also referred to as HDM2, for the human variant).[46] Small molecule disruptors have already been reported for MDM2, a protein that hetero-dimerizes with the tumor suppressor p53.[55] We initially used the Diversity Set I structures available from the National Cancer Institute for this comparative study. Although NCI Diversity Set II has superseded NCI Diversity Set I, NCI

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

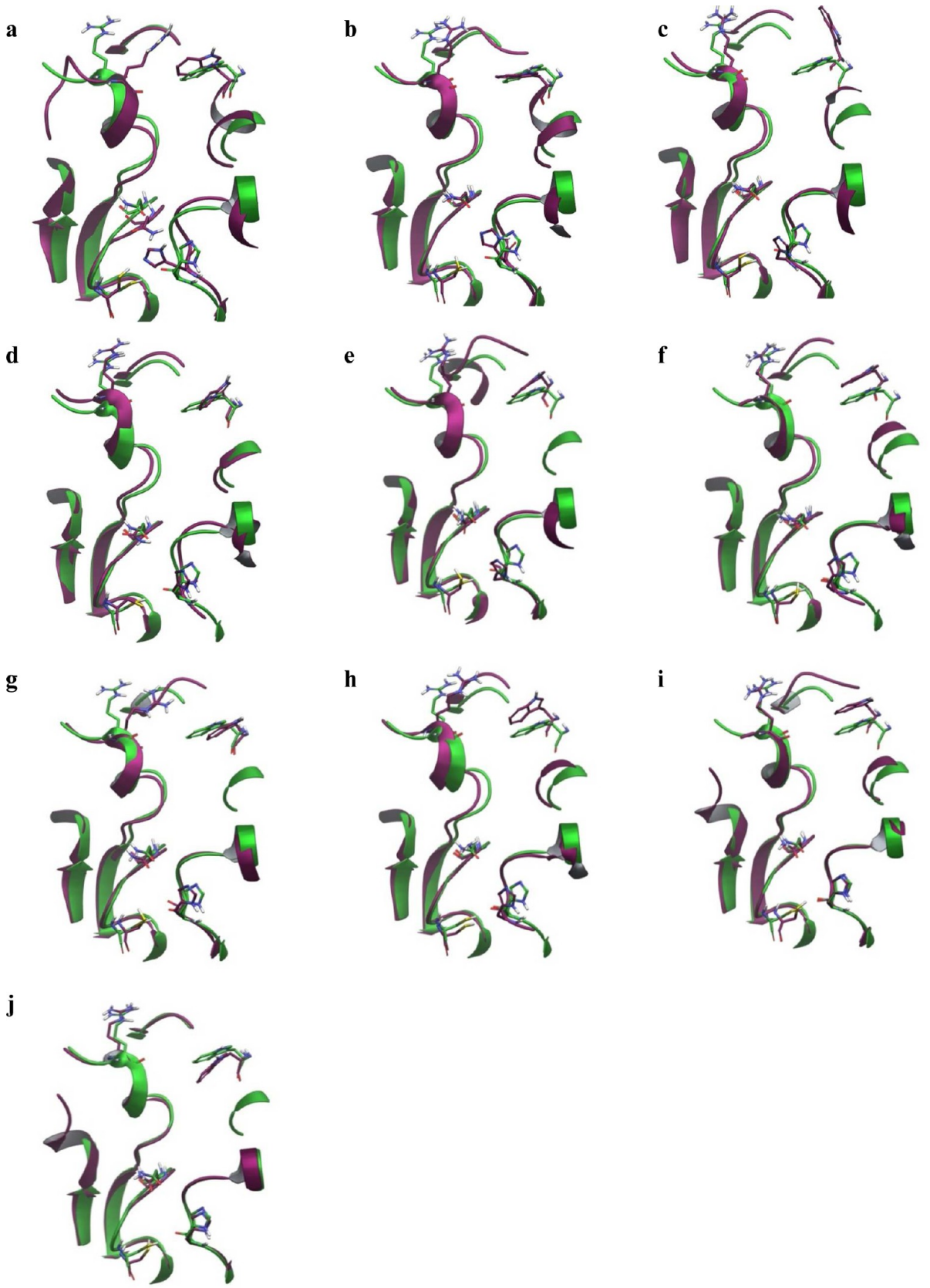


Figure 14. Conformations of the active site residues from the selected snapshots superimposed on the starting structure. (a) Frame 160 superimposed on Frame 0 (green). (b) Frame 580 superimposed on Frame 0. (c) Frame 820 superimposed on Frame 0. (d) Frame 1260 superimposed on Frame 0. (e) Frame 1700 superimposed on Frame 0. (f) Frame 2020 superimposed on Frame 0. (g) Frame 2330 superimposed on Frame 0. (h) Frame 2760 superimposed on Frame 0. (i) Frame 3420 superimposed on Frame 0. (j) Frame 4140 superimposed on Frame 0.

dx.doi.org/10.1021/ci3003528 | *J. Chem. Inf. Model.* XXXX, XXX, XXX–XXX

Table 2. SiteMap Analyses on Different Snapshots of Monomer Site 2

title	SiteScore[a]	size[b]	*Dscore*[c]	volume[d]	exposure[d]	enclosure[d]	contact[e]	hydrophobic[f]	hydrophilic[f]	don/acc[g]
Frame 0	1.01	103	1.07	288.81	0.69	0.64	0.81	0.84	0.77	0.98
Frame 160	0.65	29	0.51	56.94	0.59	0.68	0.97	0.21	1.24	0.25
Frame 580	1.04	338	1.02	828.35	0.51	0.75	1.00	0.47	1.14	0.81
Frame 820	0.70	36	0.64	73.06	0.59	0.67	0.89	0.35	0.99	0.28
Frame 1700	1.03	323	0.99	591.33	0.42	0.75	1.00	0.44	1.22	1.04
Frame 1260	1.05	343	1.08	735.39	0.51	0.75	0.94	0.78	0.97	0.86
Frame 2020	1.00	117	0.98	288.81	0.52	0.69	0.91	0.20	1.16	1.32
Frame 2330	0.99	101	0.89	262.74	0.59	0.69	0.94	0.25	1.43	0.85
Frame 2760	0.99	143	0.95	338.54	0.58	0.68	0.91	0.27	1.23	0.82
Frame 3420	0.87	74	0.82	180.08	0.69	0.66	0.86	0.19	1.19	0.82
Frame 4140	0.92	87	0.91	270.97	0.69	0.64	0.82	0.16	1.08	1.15

[a]Used to identify and compare binding sites; a score of 0.8 or higher is generally considered a binding site. [b]Number of site points for the given binding site; a size of 132 or higher is optimal for submicromolar sites. [c]Criterion to decide whether or not the site is druggable; a score greater than 0.98 is considered "druggable". [d]Refers to the openness of the binding site to the solvent; lower exposure scores and higher enclosure scores are considered better. [e]A measure of vdW non-bonded interactions of the site points with the recepto; average score for submicromolar sites being 1.0. [f]Refers to the hydrophobic and hydrophilic properties of the binding site, respectively; average scores are 1.0 for submicromolar sites. [g]Refers to the ability of the ligand to donate or accept hydrogen bonds; average score for submicromolar sites is 0.76.

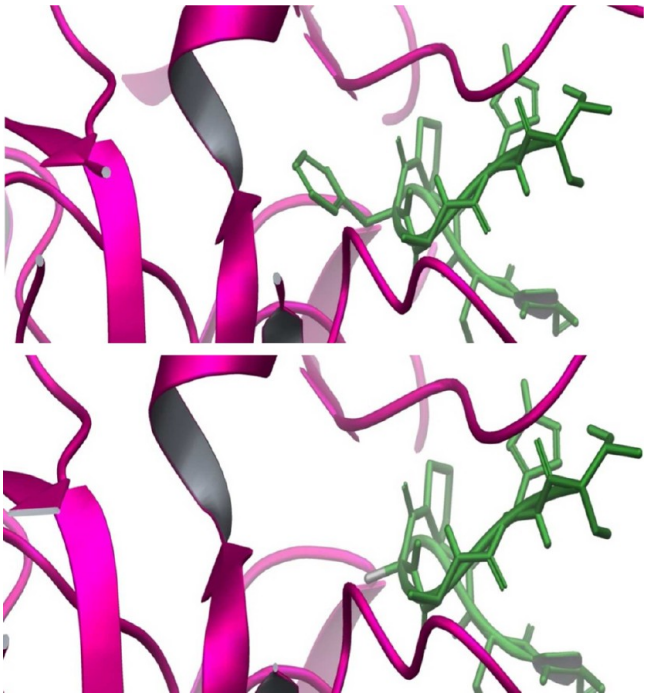


Figure 15. Truncated 7-mer peptide in the dimer interface for the WT and F87A mutant.

Table 3. Binding Free Energies from MM/GBSA on WT vs F87A Mutant

time (ns)	ΔG WT	ΔG mutant	$\Delta\Delta G$[a]
0.00	−96.47	−67.30	−29.17
0.50	−98.61	−82.81	−15.80
1.00	−103.53	−89.61	−13.92
1.50	−90.44	−90.01	−0.43
2.00	−89.94	−40.20	−49.74
2.50	−84.33	−76.40	−7.93
3.00	−80.72	−53.19	−27.53
3.50	−90.19	−74.41	−15.78
4.00	−97.65	−58.72	−38.93
4.50	−97.30	−79.22	−18.08
5.00	−88.36	−78.27	−10.09
5.50	−99.86	−66.39	−33.47
6.00	−95.85	−76.48	−19.37
Average	−93.33	−71.77	−21.56

[a]$\Delta\Delta G$ calculated as the difference in the binding free energies (MM/GBSA) between the wild type and mutant (F87A) of FabH.

Diversity Set I still encapsulates the chemical diversity of 140,000 compounds in NCI's Plated Set, which in turn represents well over a quarter of a million compounds. The structures from the NCI Diversity Set I were prepared using Ligprep[32] as previously described.[32] The prepared ligand set was then docked to the receptors using GLIDE SP docking, and the best scoring pose for each compound was selected using a script available from Schrödinger.

Virtual Screening. Comparative docking studies were performed using Monomer-Site2 (the dimer interface region), the MDM2 (PDB ID: 2AXI) receptor, the biological active site (Monomer-Site1), and Dimer-site5 (shallow binding region in the dimer interface). Although Dimer-Site5 was identified as a potential binding region in the dimer, only the monomer was used as the receptor for docking studies because we were interested in inhibiting the dimer formation. The dimer interface (Monomer-Site2) and MDM2 (PDB ID: 2AXI) had a very similar distribution of scores (Figure 16). The shallow binding region at the dimer interface region (Dimer-Site5) had poor docking scores as expected because it does not have a

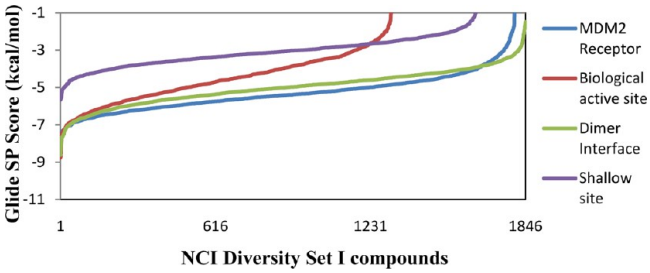


Figure 16. Distribution of docking (GLIDE-SP) scores from docking the NCI Diversity Set I to MDM2 receptor, biological active site of FabH (Monomer-Site1), dimer interface region of FabH (Monomer-Site2), and a shallow binding region at the dimer interface (Dimer-Site5).

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

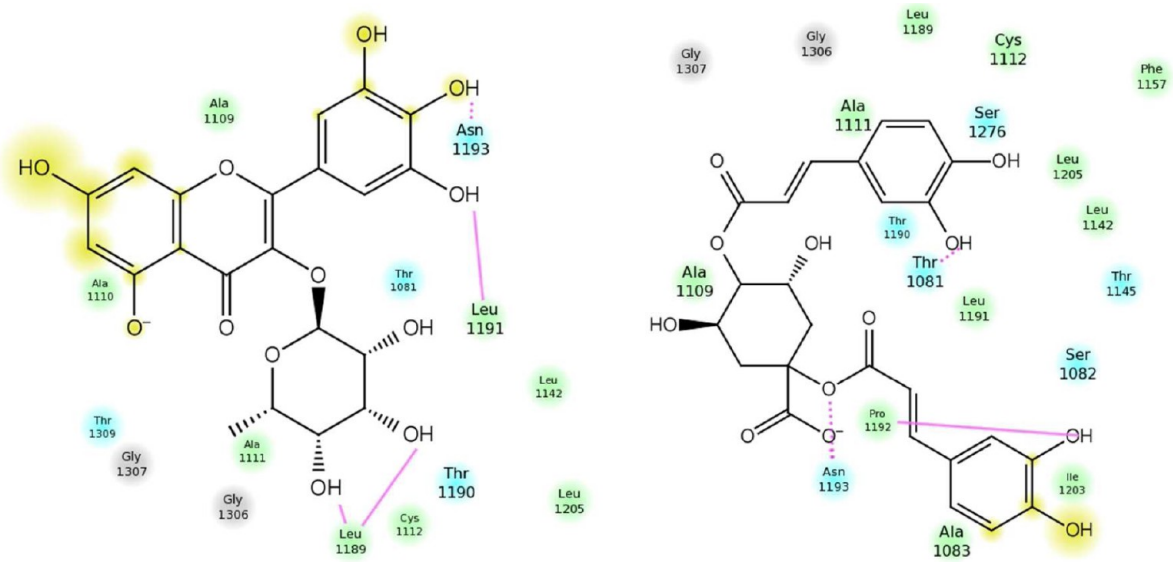


Figure 17. Ligand interaction diagram of compounds **NSC19803** (left) and **NSC91529** (right) in the dimer interface binding site. Residues are named using their three-letter amino acid code. Hydrophobic residues are colored green; polar residues are colored light purple; basic residues are colored blue; acidic residues are colored red; hydrogen bonds are drawn with a pink line.

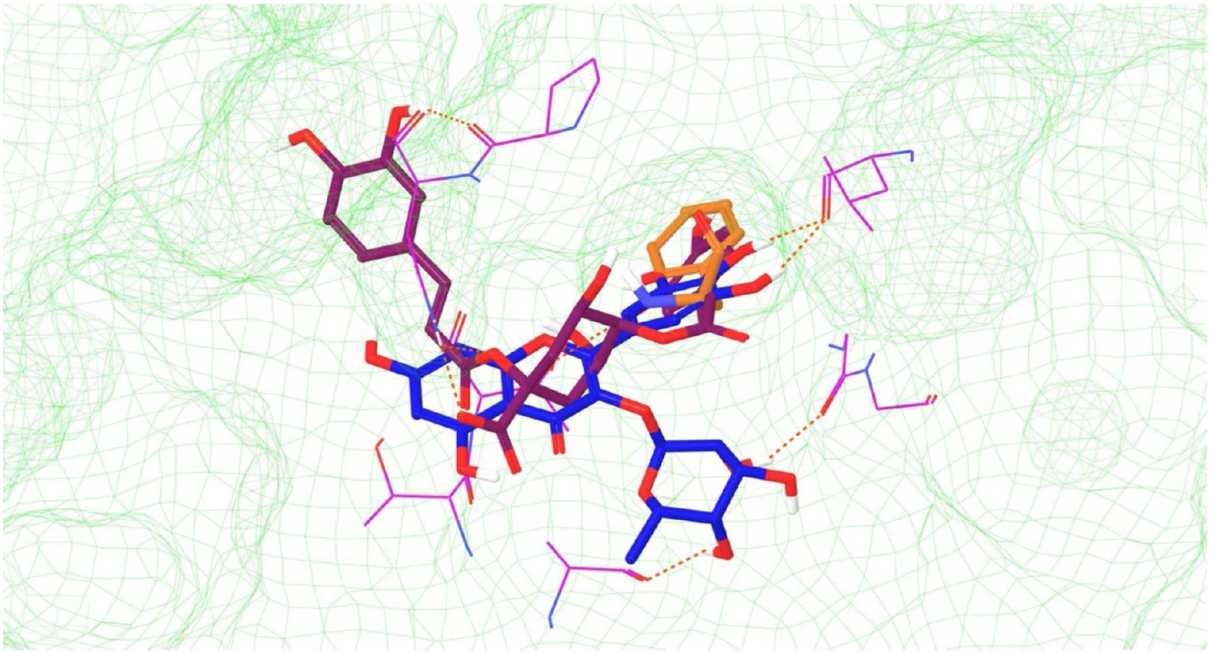


Figure 18. Compounds **NSC91529** (purple) and **NSC19803** (blue) bound to FabH monomer at the interface region (surface shown as mesh surface in green). Phe87 from the other monomer is shown in rust-orange.

well-defined pocket. The biological active site (Monomer-Site1) on the other hand exhibited better docking scores than the shallow binding site (Dimer-Site5), but it showed poorer docking scores than either PPI sites (MDM2 receptor and Monomer-Site2). This suggests that induced-fit effects may have to be included for a more accurate docking of inhibitors at the biological active site. These data correlate well with the previously described analyses using SiteMap interpreting that Monomer-Site2 at the dimer interface region is a potential druggable binding site.

Docking studies with the NCI Diversity Set II were then performed on the starting structure of the monomer, which corresponds to Frame 0 and also representative members of each cluster from the MD simulation: frames 160, 580, 820, 1260, 1700, 2020, 2330, 2640, 3420, and 4140 corresponding to simulation times of 0.77, 2.78, 3.94, 6.05, 8.16, 9.70, 11.18, 13.25, 16.42, and 19.87 ns, respectively. Docking was done using the compounds in the NCI Diversity Set II after Ligprep[32] was performed on the ligand set. The compounds were screened initially using GLIDE standard precision docking (SP) and then using extra precision (XP) docking. The poses of the ligands were sorted using the best GLIDE XP scores for each ligand. Docking studies on the starting structure of the monomer revealed that compounds **NSC91529** and **NSC19803** from the NCI Diversity Set II docked best in the dimer interface region with docking scores of −11.75 and

dx.doi.org/10.1021/ci3003528 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

−10.20 kcal/mol, respectively, using the XP docking. The ligand interaction energy diagrams showing the interaction of the ligands, **NSC91529** and **NSC19803**, with the residues of the receptor are shown in Figure 17.

Figure 18 shows the docked pose of the ligands, **NSC91529** and **NSC19803** in the dimer interface region of FabH. The phenyl ring of **NSC91529** and pyranose ring (monosaccharide unit) of **NSC19803** were found to occupy the hydrophobic pocket formed by Phe1087 of the other monomer. Interestingly, in the second lowest energy pose for **NSC19803** (docking score: −10.05 kcal/mol), the trihydroxyphenyl substituent occupied the Phe87 binding pocket.

Representative members of the top clusters as found by DESMOND were also used for docking studies because the monomer would not retain the original conformation in the absence of its dimeric partner. Nonetheless, all the snapshots showed very high affinity for compounds **NSC91529** and/or **NSC19803** with docking scores ranging between −9.18 and −9.78 kcal/mol. The binding modes for both compounds were similar for each representative member of the cluster. The receptors had an RMSD that ranged from 1.65−3.34 Å with respect to Frame 0, and compound **NSC19803** in the different snapshots had an RMSD ranging from 7.03−13.2 Å with respect to the starting structure.

CONCLUSION

In conclusion, we have shown that the dimer interface in FabH is potentially a viable target for designing a new class of inhibitors. Although disrupting protein−protein interactions is still generally considered to be a difficult task, the recent emergence of PPI disruptors suggests the possibility of PPI disruption by small molecules or peptides that target the "hot spots" in the PPI. The MD simulations we have described strongly suggest that FabH monomers are stabilized through dimerization and that the dimeric state is essential for enzymatic activity. Solvated monomers are most likely not capable of binding the substrate at the biological active site due to large conformational changes occurring in the solvated monomer relative to its dimeric state that lead to a significant change in the 3-D architecture of the active site. Any change in conformation at the interface region will most likely alter the shape of the long cylindrical tunnel-like active site, which in turn would not accommodate substrate binding, rendering the enzyme inactive. Comparative docking score analysis of the NCI Diversity Set I with the FabH dimer interface, a shallow site at the dimer interface, the FabH catalytic site, and the MDM2 receptor (for which inhibitors have been previously discovered that disrupt its association with its native protein partner p53) further suggests that the dimer interface could be an interesting target for the discovery of small molecules that inhibit dimerization. This comparative study also demonstrates that the dimer interface in FabH behaves very similar to known protein−protein interactions. Upon virtual screening of compounds from the NCI Diversity Set II, we have identified compounds **NSC91529** and **NSC19803** as potential ligands capable of binding at the dimer interface region with good affinity. Do they bind with sufficient affinity to inhibit dimerization and exhibit biochemical/physiological relevance? That question will have to await experimental studies.

ASSOCIATED CONTENT

Supporting Information

Torsion angle fluctuations of residues in the FabH dimer interface region, histograms of GLIDE scores for the top-ranking NCI Diversity Set I compounds docked to FabH, conformations of residues complementary to Phe87 from the selected trajectory snapshots, sequence alignment for bacterial FabH species, additional PCA data and methodology used, and the GLIDE scoring functions used. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: wguida@usf.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the National Cancer Institute for making their databases publicly available. We thank The University of South Florida for start-up funding.

REFERENCES

(1) Infectious Diseases Report. WHO.http://www.who.int/infectious-disease-report/pages/ch1text.html (accessed 03/26/2013).

(2) Heath, R. J.; Rock, C. O. The Claisen condensation in biology. *Nat. Prod. Rep.* **2002**, *19*, 581−596.

(3) Zhang, Y.-M.; White, S. W.; Rock, C. O. Inhibiting bacterial fatty acid synthesis. *J. Biol. Chem.* **2006**, *281*, 17541−17544.

(4) Khandekar, S. S.; Gentry, D. R.; Van Aller, G. S.; Warren, P.; Xiang, H.; Silverman, C.; Doyle, M. L.; Chambers, P. A.; Konstantinidis, A. K.; Brandt, M.; Daines, R. A.; Lonsdale, J. T. Identification, substrate specificity, and inhibition of the *Streptococcus pneumoniae* β-ketoacyl-acyl carrier protein synthase III (FabH). *J. Biol. Chem.* **2001**, *276*, 30024−30030.

(5) Parsons, J. B.; Frank, M. W.; Subramanian, C.; Saenkham, P.; Rock, C. O. Metabolic basis for the differential susceptibility of Gram-positive pathogens to fatty acid synthesis inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 15378−15383.

(6) Daines, R. A.; Pendrak, I.; Sham, K.; Van Aller, G. S.; Konstantinidis, A. K.; Lonsdale, J. T.; Janson, C. A.; Qiu, X.; Brandt, M.; Khandekar, S. S.; Silverman, C.; Head, M. S. First X-ray cocrystal structure of a bacterial FabH condensing enzyme and a small molecule inhibitor achieved using rational design and homology modeling. *J. Med. Chem.* **2002**, *46*, 5−8.

(7) Li, H.-Q.; Luo, Y.; Lv, P.-C.; Shi, L.; Liu, C.-H.; Zhu, H.-L. Design and synthesis of novel deoxybenzoin derivatives as FabH inhibitors and anti-inflammatory agents. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 2025−2028.

(8) Wang, J.; Kodali, S.; Lee, S. H.; Galgoci, A.; Painter, R.; Dorso, K.; Racine, F.; Motyl, M.; Hernandez, L.; Tinney, E.; Colletti, S. L.; Herath, K.; Cummings, R.; Salazar, O.; González, I.; Basilio, A.; Vicente, F.; Genilloud, O.; Pelaez, F.; Jayasuriya, H.; Young, K.; Cully, D. F.; Singh, S. B. Discovery of platencin, a dual FabF and FabH inhibitor with in vivo antibiotic properties. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7612−7616.

(9) Young, K.; Jayasuriya, H.; Ondeyka, J. G.; Herath, K.; Zhang, C.; Kodali, S.; Galgoci, A.; Painter, R.; Brown-Driver, V.; Yamamoto, R.; Silver, L. L.; Zheng, Y.; Ventura, J. I.; Sigmund, J.; Ha, S.; Basilio, A.; Vicente, F.; Tormo, J. R.; Pelaez, F.; Youngman, P.; Cully, D.; Barrett, J. F.; Schmatz, D.; Singh, S. B.; Wang, J. Discovery of FabH/FabF inhibitors from natural products. *Antimicrob. Agents Chemother.* **2006**, *50*, 519−26.

(10) Chan, D. I.; Vogel, H. J. Current understanding of fatty acid biosynthesis and the acyl carrier protein. *Biochem. J.* **2010**, *430*, 1−19.

(11) Alhamadsheh, M. M.; Musayev, F.; Komissarov, A. A.; Sachdeva, S.; Wright, H. T.; Scarsdale, N.; Florova, G.; Reynolds, K. A. Alkyl-CoA disulfides as inhibitors and mechanistic probes for FabH enzymes. *Chem. Biol.* **2007**, *14*, 513−524.
(12) Alhamadsheh, M. M.; Waters, N. C.; Huddler, D. P.; Kreishman-Deitrick, M.; Florova, G.; Reynolds, K. A. Synthesis and biological evaluation of thiazolidine-2-one 1,1-dioxide as inhibitors of Escherichia coli beta-ketoacyl-ACP-synthase III (FabH). *Bioorg. Med. Chem. Lett.* **2007**, *17*, 879−883.
(13) Alhamadsheh, M. M.; Waters, N. C.; Sachdeva, S.; Lee, P.; Reynolds, K. A. Synthesis and biological evaluation of novel sulfonyl-naphthalene-1,4-diols as FabH inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 6402−6405.
(14) Heath, R. J.; White, S. W.; Rock, C. O. Inhibitors of fatty acid synthesis as antimicrobial chemotherapeutics. *Appl. Microbiol. Biotechnol.* **2002**, *58*, 695−703.
(15) Han, L.; Lobo, S.; Reynolds, K. A. Characterization of beta-ketoacyl-acyl carrier protein synthase III from *Streptomyces glaucescens* and its role in initiation of fatty acid biosynthesis. *J. Bacteriol.* **1998**, *180*, 4481−4486.
(16) Khandekar, S. S.; Daines, R. A.; Lonsdale, J. T. Bacterial beta-ketoacyl-acyl carrier protein synthases as targets for antibacterial agents. *Curr. Protein Pept. Sci.* **2003**, *4*, 21−29.
(17) Qiu, X.; Janson, C. A.; Konstantinidis, A. K.; Nwagwu, S.; Silverman, C.; Smith, W. W.; Khandekar, S.; Lonsdale, J.; Abdel-Meguid, S. S. Crystal structure of β-ketoacyl-acyl carrier protein synthase III. *J. Biol. Chem.* **1999**, *274*, 36465−36471.
(18) Qiu, X.; Janson, C. A.; Smith, W. W.; Head, M.; Lonsdale, J.; Konstantinidis, A. K. Refined structures of β-ketoacyl-acyl carrier protein synthase III. *J. Mol. Biol.* **2001**, *307*, 341−356.
(19) Konstantinidis, A. K.; Lonsdale, J.; Timothy, V. A.; Glenn, S. Methods of Modulating FabH Activity. U.S. patent application PCT/US00/12250, U.S. Patent WO 00/67780,SmithKline Beecham Corporation: Philadelphia, PA, November 16, 2000.
(20) Khandekar, S. S.; Konstantinidis, A. K.; Silverman, C.; Janson, C. A.; McNulty, D. E.; Nwagwu, S.; Van Aller, G. S.; Doyle, M. L.; Qiu, X.; Lonsdale, J. Expression, purification and crystallization of the *Escherichia coli* selenomethionyl beta-ketoacyl-acyl carrier protein synthase III. *FASEB J.* **2000**, *14*, A1322−A1322.
(21) Pérez-Castillo, Y.; Froeyen, M.; Cabrera-Pérez, M.; Nowé, A. Molecular dynamics and docking simulations as a proof of high flexibility in *E. coli* FabH and its relevance for accurate inhibitor modeling. *J. Comput.-Aided Mol. Des.* **2011**, *25*, 371−393.
(22) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein−protein interfaces. *Nature* **2007**, *450*, 1001−1009.
(23) Zutshi, R.; Brickner, M.; Chmielewski, J. Inhibiting the assembly of protein−protein interfaces. *Curr. Opin. Chem. Biol.* **1998**, *2*, 62−66.
(24) He, X.; Reynolds, K. A. Purification, characterization, and identification of novel inhibitors of the β-ketoacyl-acyl carrier protein synthase III (FabH) from *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **2002**, *46*, 1310−1318.
(25) Pieniazek, D.; Rayfield, M.; Hu, D. J.; Nkengasong, J.; Wiktor, S. Z.; Downing, R.; Biryahwaho, B.; Mastro, T.; Tanuri, A.; Soriano, V.; Lal, R.; Dondero, T.; Grp, H. I. V. V. W. Protease sequences from HIV-1 group M subtypes A−H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naive individuals worldwide. *Aids* **2000**, *14*, 489−495.
(26) *Maestro*, version 9.2; Schrödinger, Inc.: New York, 2011.
(27) *Schrödinger Suite 2011, Protein Preparation Wizard*, Epik version 2.2; Schrödinger, Inc.: New York, 2011.
(28) *Impact*, version 5.7; Schrödinger, Inc.: New York, 2011.
(29) *Prime*, version 3.0; Schrödinger, Inc.: New York, 2011.
(30) *Desmond Molecular Dynamics System*, version 3.0; D. E. Shaw Research: New York, 2011.
(31) Schrödinger, Inc.: New York, 2011.
(32) *LigPrep*, version 2.5; Schrödinger, Inc.: New York, 2011.
(33) *Epik*, version 2.2; Schrödinger, Inc.: New York, 2011.
(34) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A software program for pK (a) prediction and protonation state generation for drug-like molecules. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681−691.
(35) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(36) Halgren, T. New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* **2007**, *69*, 146−148.
(37) Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377−389.
(38) Jorgensen, W.; Maxwell, D.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(39) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2010**, *51*, 69−82.
(40) Sippel, M.; Sotriffer, C. A. Molecular dynamics simulations of the HIV-1 integrase dimerization interface: Guidelines for the design of a novel class of integrase inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 604−614.
(41) *SiteMap*, version 2.5; Schrödinger, Inc.: New York, 2011.
(42) Liu, Q.; Li, J. Protein binding hot spots and the residue−-residue pairing preference: A water exclusion perspective. *BMC Bioinf.* **2010**, *11*, 244−258.
(43) Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C. Binding of small molecules to an adaptive protein−protein interface. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1603−1608.
(44) Vigers, G. P. A.; Dripps, D. J.; Edwards, C. K.; Brandhuber, B. J. X-ray crystal structure of a small antagonist peptide bound to interleukin-1 receptor type 1. *J. Biol. Chem.* **2000**, *275*, 36927−36933.
(45) Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E. A. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **2004**, *303*, 844−848.
(46) Fasan, R.; Dias, R. L. A.; Moehle, K.; Zerbe, O.; Obrecht, D.; Mittl, P. R. E.; Grütter, M. G.; Robinson, J. A. Structure−activity studies in a family of β-hairpin protein epitope mimetic inhibitors of the p53−HDM2 protein−protein interaction. *ChemBioChem* **2006**, *7*, 515−526.
(47) Degterev, A.; Lugovskoy, A.; Cardone, M.; Mulley, B.; Wagner, G.; Mitchison, T.; Yuan, J. Identification of small-molecule inhibitors of interaction between the BH3 domain and Bcl-xL. *Nat. Cell Biol.* **2001**, *3*, 173−182.
(48) Lee, E. F.; Czabotar, P. E.; Smith, B. J.; Deshayes, K.; Zobel, K.; Colman, P. M.; Fairlie, W. D. Crystal structure of ABT-737 complexed with Bcl-xL: Implications for selectivity of antagonists of the Bcl-2 family. *Cell Death Differ.* **2007**, *14*, 1711−1713.
(49) Qiu, X.; Choudhry, A. E.; Janson, C. A.; Grooms, M.; Daines, R. A.; Lonsdale, J. T.; Khandekar, S. S. Crystal structure and substrate specificity of the β-ketoacyl-acyl carrier protein synthase III (FabH) from Staphylococcus aureus. *Protein Sci.* **2005**, *14*, 2087−2094.
(50) Yang, L.-W.; Bahar, I. Coupling between catalytic site and collective dynamics: A requirement for mechanochemical activity of enzymes. *Structure* **2005**, *13*, 893−904.
(51) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, R Development Core Team, 2008. ISBN 3-900051-07-0, http://www.R-project.org (accesed 03/26/20013).
(52) The R FAQ, Kurt Hornik, 2012, ISBN 3-900051-08-9,http://CRAN.R-project.org/doc/FAQ/R-FAQ.html (accessed 03/26/2013).
(53) Modified from a script originally written by Jordan Monnet and mentioned by Niel Henriksen in the public forum: RE: [AMBER]

How To Obtain Dialsplot. http://archive.ambermd.org/201005/0334.html (accesed 03/26/20013).

(54) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Computational alanine scanning mutagenesis: An improved methodological approach. *J. Comput. Chem.* **2007**, *28*, 644−654.

(55) Lawrence, H. R.; Li, Z.; Richard Yip, M. L.; Sung, S.-S.; Lawrence, N. J.; McLaughlin, M. L.; McManus, G. J.; Zaworotko, M. J.; Sebti, S. M.; Chen, J.; Guida, W. C. Identification of a disruptor of the MDM2-p53 protein−protein interaction facilitated by high-throughput *in silico* docking. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3756−3759.