Computational Prediction of the Mode of Binding of Antitumor Lankacidin C to Tubulin

Ahmed Taha Ayoub,[*,†] Mohamed Ali Elrefaiy,[‡] and Kenji Arakawa[§]

[†]Medicinal Chemistry Department, Heliopolis University, 3 Cairo-Belbeis Desert Road, El-Nahda, Qism El-Salam, Cairo 11777, Egypt

[‡]Center of X-ray Determination for Structure of Matter (CXDS), Zewail City of Science and Technology, 6th of October City, Giza 12588, Egypt

[§]Department of Molecular Biotechnology, Graduate School of Advanced Sciences of Matter, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima City, Hiroshima 739-8530, Japan

S Supporting Information

ABSTRACT: Lankacidin C, which is an antibiotic produced by the organism *Streptomyces rochei*, shows considerable antitumor activity. The mechanism of its antitumor activity remained elusive for decades until it was recently shown to overstabilize microtubules by binding at the taxol binding site of tubulin, causing mitotic arrest followed by apoptosis. However, the exact binding mode of lankacidin C inside the tubulin binding pocket remains unknown, an issue that impedes proper structure-based design, modification, and optimization of the drug. Here, we have used computational methods to predict the most likely binding mode of lankacidin C to tubulin. We employed ensemble-based docking in different software packages, supplemented with molecular dynamics simulation and subsequent binding-energy prediction. The molecular dynamics simulations performed on lankacidin C were collectively 1.1 μs long. Also, a multiple-trajectory approach was performed to assess the stability of different potential binding modes. The identified binding mode could serve as an ideal starting point for structural modification and optimization of lankacidin C to enhance its affinity to the tubulin binding site and therefore improve its antitumor activity.
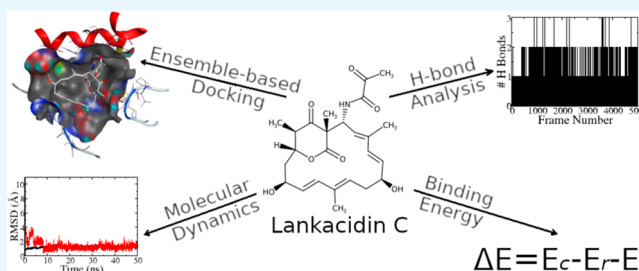
1. INTRODUCTION

Lankacidin C is the parent of a group of antibiotics (T2636) that are produced by the soil-dwelling bacteria *Streptomyces rochei*. This group of antibiotics was first isolated in 1969 by Harada and co-workers,[1] and full characterization of the group was achieved a few years later.[2−5] Lankacidin C, Figure 1a, possesses the structure of a 17-membered macrocyclic polyketide and was found to have strong antimicrobial activity against various Gram-positive bacteria, some of which are resistant to macrolide antibiotics.[6] The antimicrobial activity of lankacidin C was found to be due to interference with peptide bond formation during bacterial protein synthesis by binding at the peptidyl transferase center of the eubacterial large ribosomal subunit.[7,8] Interest in lankacidin C has grown recently, and several studies have been conducted with the aim of facilitating its synthesis and large-scale production.[9−11]

Lankacidin C and various derivatives also displayed considerable in vivo antitumor activity against various tumor models such as 6C3 HED/OG lymphosarcoma, L1210 leukemia, and B16 melanoma,[12,13] in addition to T47D breast cancer cell lines.[14] However, the mechanism of lankacidin C antitumor activity remained unclear and was usually confused with its mechanism of antimicrobial activity, until we recently identified the mechanism. The mechanism of action was predicted based on some computational simulations involving compounds structurally very similar to lankacidin C.[15] In our recent work,[14] we have shown that the mechanism of lankacidin C antitumor activity is due to binding at the taxol binding site of tubulin, which causes overstabilization of cellular microtubules constituting the mitotic spindle. This freezes the dividing cell at the metaphase, resulting in mitotic arrest followed by apoptosis.
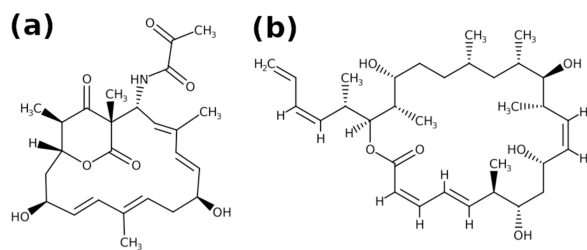

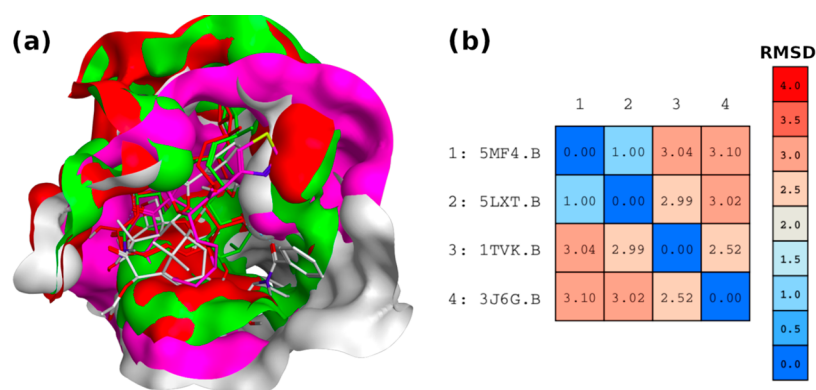**Figure 1.** Structure of (a) lankacidin C and (b) dictyostatin.

Figure 2. Taxol binding site of tubulin shows a great deal of induced fit, posing a real challenge in docking of new ligands. (a) Binding pocket changes shape to accommodate a variety of structurally diverse ligands, including paclitaxel (white, PDB ID: 3J6G[16]), discodermolide (green, PDB ID: 5LXT[17]), epothilone A (purple, PDB ID: 1TVK[18]), and dictyostatin (red, PDB ID: 5MF4[19]). (b) RMSD matrix of the binding pocket residues of the four different aforementioned tubulin PDB structures in angstrom.

Table 1. Different Representative Binding Modes for Lankacidin C and Dictyostatin Obtained from GOLD and FRED Docking Steps[a]

lankacidin C				dictyostatin			
mode	receptor	rank	score	mode	receptor	rank	score
GOLD1	1TVK	1	65.34	GOLD1	5MF4	1	70.04
GOLD3	1TVK	3	61.86	GOLD2	5LXT	2	69.40
GOLD6	1TVK	6	59.84	GOLD3	5LXT	3	68.88
GOLD12	1TVK	12	58.63	GOLD4	1TVK	4	68.76
FRED1	5LXT	1	−9.20	FRED1	5LXT	1	−6.95
FRED2	1TVK	2	−8.57	FRED2	1TVK	2	−6.76
FRED3	5MF4	3	−8.50	FRED3	3J6G	3	−6.46
FRED4	3J6G	4	−5.45	FRED4	5MF4	4	−5.67

[a]The score in the case of GOLD is based on piecewise linear potential fitness and in the case of FRED is based on Chemgauss4 scoring function.

Despite identifying the antitumor mechanism of action and the tubulin binding site of lankacidin C, it is still unknown how lankacidin C binds at the taxol binding site. X-ray crystallographic studies managed to identify the binding mode of lankacidin C to the eubacterial large ribosomal subunit associated with its antimicrobial activity.[7,8] However, the binding mode of lankacidin C to the taxol binding site of tubulin, associated with its antitumor activity, remains unknown. We have previously done some computational simulations on lankacidin C and related analogues, and some preliminary data on the binding mode were available.[14,15] Nevertheless, a comprehensive study on the binding mode of lankacidin C to tubulin was never performed. In this study, we utilize cutting-edge computational techniques to identify the binding mode of lankacidin C in the taxol binding site of tubulin. We employ several ensemble-based docking experiments, utilizing different computer software, to identify several potential binding modes. The study is then supplemented with single and parallel molecular dynamics (MD) simulations (totaling 1.1 μs) and binding-energy calculations to assess the various predicted binding modes and identify the most likely one. The ensemble-based docking protocol is capable of partially including receptor flexibility, whereas molecular dynamics simulations can thereafter allow full receptor flexibility. This approach was necessary since the taxol binding site shows a great deal of induced fit and the binding pocket can take several different shapes to accommodate a variety of structurally diverse drugs, such as paclitaxel, discodermolide, epothilone A, and dictyostatin. Figure 2 shows the different

shapes adopted by the taxol binding pocket while hosting those different ligands. It also shows a root-mean-square deviation (RMSD) matrix quantifying the structural differences between the shapes of the binding pockets in each case. Excluding the decent similarity between the binding pockets of 5MF4 and 5LXT (RMSD < 2.0 Å), the shapes of the binding pockets are significantly different with RMSD values >2.0 Å. Furthermore, we applied the same protocol outlined in this study to predict the binding mode of dictyostatin (Figure 1b), whose experimental binding mode is available through X-ray crystallography (Protein Data Bank (PDB) ID: 5MF4). We used this as a proof of concept to validate our protocol.

Finding out the correct binding pose of lankacidin C in the taxol binding site should be of great importance and shall serve as an excellent starting point for structure-based design of lankacidin C. After the correct binding mode of lankacidin C is established, the interactions inside the binding pocket can be thoroughly studied and informed decisions regarding structural modifications can be made. Such modifications can be made to optimally exploit the binding cavity on the road to tailor-design derivatives that should have better antitumor activity than the parent lankacidin C.

2. RESULTS AND DISCUSSION

2.1. Ensemble-Based Docking. In this part, we docked different lankacidin C conformers into a set of four different conformations of the taxol binding pocket of tubulin using GOLD and FRED docking software. Tubulin structures were obtained from 1TVK (tubulin bound to epothilone A), 5MF4
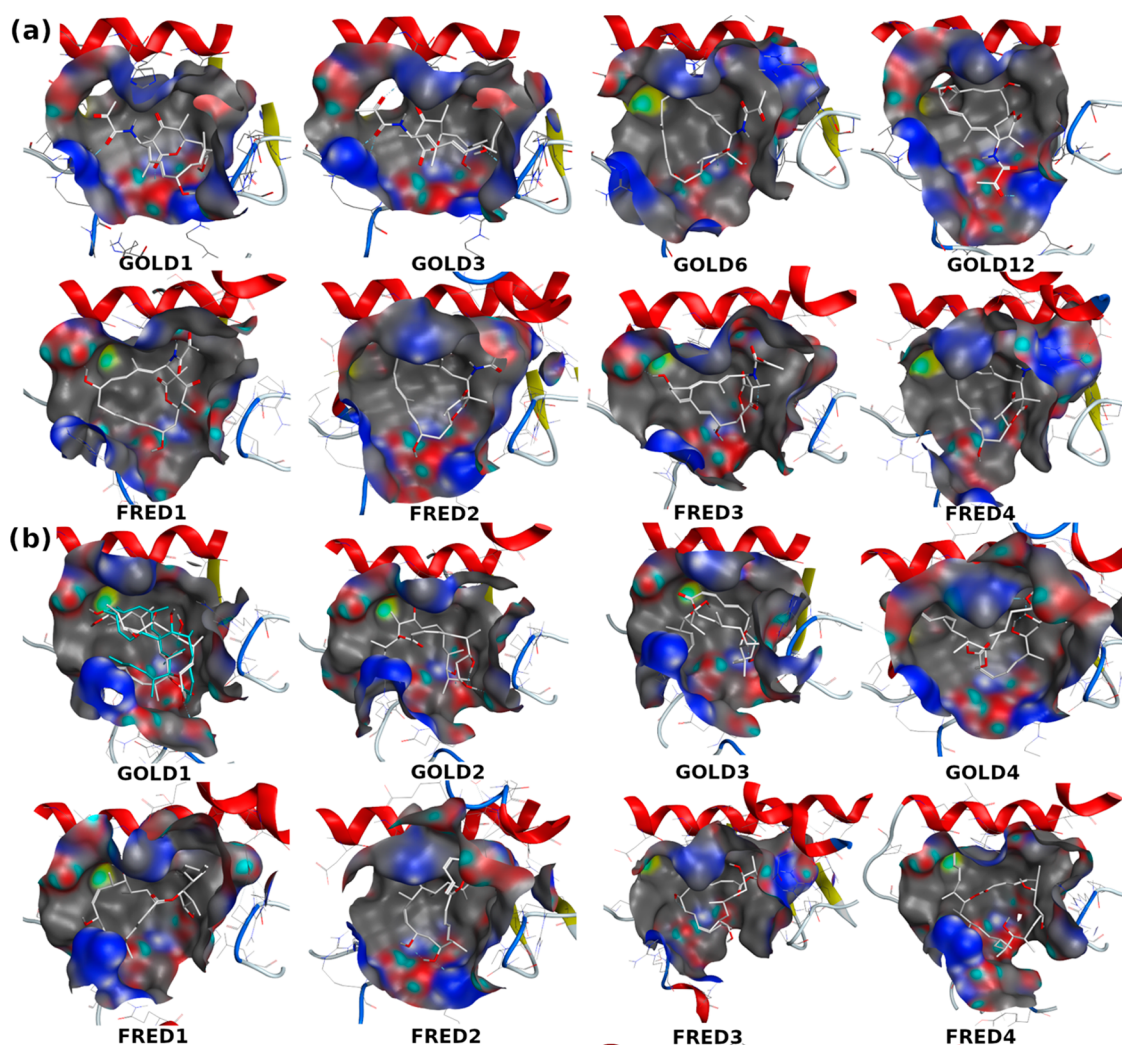
Figure 3. Eight different binding modes selected through docking for (a) lankacidin C and (b) dictyostatin. The native binding mode for dictyostatin is shown together with the GOLD1 binding mode in (a), colored in cyan.
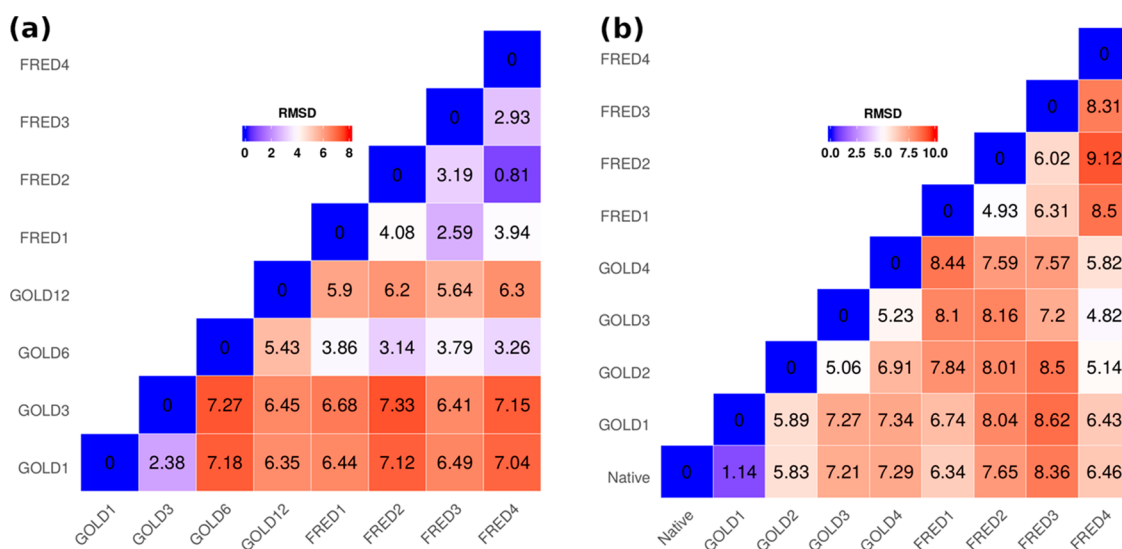

Figure 4. Heavy-atom RMSD matrix quantifying, in angstrom, the differences between the eight different binding poses after fitting the atoms of the binding pocket in the case of (a) lankacidin C and (b) dictyostatin.

(tubulin bound to dictyostatin), 5LXT (tubulin bound to discodermolide) and 3J6G (tubulin bound to paclitaxel), all of which are available on the Protein Data Bank. This helps capture the binding site flexibility and account for induced fit.

The ensemble-based docking returned a list of potential binding modes that were ranked according to their scores, as shown in Table 1. We selected a set of representative binding modes from each docking protocol. In the case of GOLD docking, the results of the ensemble-based docking from all protein conformations were compiled together and the best-ranked docking poses were visually compared. The poses were narrowed down to the first four unique binding modes. In the case of lankacidin C, those four unique binding modes were named GOLD1, GOLD3, GOLD6, and GOLD12, each of which represents a specific lankacidin C conformation and orientation inside a specific protein conformation. These binding modes are numbered according to their scoring ranks. It was found that all of these poses came from the protein conformation in 1TVK. In the case of FRED docking, each of the four docking runs was treated separately. The poses from each run, i.e., against each protein conformation, were ranked separately, and the best-scoring pose was selected. The reason we did this is that GOLD docking returned all of the best results from the 1TVK protein conformation. Therefore, in the FRED docking, we wanted to make sure that we take the best pose from each of the different protein conformations to make sure we cover all of the available protein conformational space in our analysis. The four binding modes resulting from FRED run were ranked according to their Chemgauss4 score as listed in Table 1. From the results of docking, the binding modes GOLD1 and FRED1 are the highest-ranked binding modes according to the two different docking algorithms. Therefore, one of these two modes is likely to be the correct binding mode. This hypothesis will be tested in the next sections. Figure 3a shows each of the eight different binding modes identified in the ensemble-based docking runs. The figure shows how the orientation of the ligand can sometimes be completely different as some of the poses are 180° flipped with respect to the others. Figure 4a shows a matrix of the RMSD of the ligand heavy atoms between the eight different binding modes after fitting the atoms in the binding pocket. The matrix shows that some modes are similar, whereas others are completely different. Most of the modes obtained from GOLD are significantly different with many RMSD values significantly greater than 2.0 Å. This is expected since these modes were carefully selected from the top-scoring poses to maximize diversity and uniqueness. Other modes, especially those obtained from FRED, are relatively more similar since pose diversity was not the factor in choosing them, rather receptor conformation diversity. Examining the results of the ensemble-based docking of dictyostatin presented in Table 1, we can see that GOLD and FRED protocols similarly returned four binding modes each. The best mode in GOLD docking, GOLD1, was bound to the receptor conformation coming from 5MF4, whereas in the case of FRED docking, it came from the receptor conformation in 5LXT. Since the native crystal structure of dictyostatin is indeed found in 5MF4, the GOLD docking protocol seems to have performed better this time. The binding modes are all shown in Figure 3b, and an RMSD matrix is shown in Figure 4b. Comparing all of the binding modes to the native crystal structure of dictyostatin bound to tubulin, it is clear that the GOLD1 binding mode is the closest one to the native structure. The GOLD1 binding mode is shown overlaid to the native crystal structure in Figure 3b. On the other hand, the RMSD matrix in Figure 4b shows that GOLD1 is only 1.14 Å different from the native binding mode, whereas all other binding modes are at least 4.82 Å

different. In short, these findings manifest the ability of our docking protocol to find the correct binding mode of dictyostatin.

Despite the good performance of the docking protocol in the case of dictyostatin, most of the time the results of docking and scoring are questionable even when ensemble-based docking is performed. This is because to assess the binding mode correctly, both the ligand and the protein have to be completely free to explore different conformations in the bound state. In addition, the binding energy (BE) between the two entities should be calculated as an ensemble average or a time average taking into account all of the conformations explored by the bound ligand–protein complex in comparison to the free ligand and protein. Equation 1 shows how the ensemble average of a property A, in our case it is the energy, which is a function of positions ($\mathbf{q}$) and momenta ($\mathbf{p}$), is calculated. In this case, the Boltzmann factor ($e^{-E(\mathbf{p},\mathbf{q})/kT}$) is used as a weighting factor, where E is the energy of the microstate, k is the Boltzmann constant, and T is the temperature in kelvin. The integration is done over the entire phase space.

$$\langle \mathbf{A} \rangle_{\text{ensemble}} = \frac{\iint \mathbf{A}(\mathbf{p},\, \mathbf{q})\, e^{-E(\mathbf{p},\mathbf{q})/kT}\, d\mathbf{p}\, d\mathbf{q}}{\iint e^{-E(\mathbf{p},\mathbf{q})/kT}\, d\mathbf{p}\, d\mathbf{q}} \tag{1}$$

The time average of property A can also be calculated according to eq 2, where the positions and momenta are now functions of time, which extends from zero to ∞.

$$\langle \mathbf{A} \rangle_{\text{time}} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} \mathbf{A}(\mathbf{p}(t),\, \mathbf{q}(t))\, dt \tag{2}$$

The ergodic hypothesis maintains that in ergodic systems, the ensemble average is equal to the time average as demonstrated in eq 3. This is where molecular dynamics comes into play where the system is simulated for a finite period of time hoping that this simulation will capture the most important micro-states of the system. After that, a representative set of microstates M (i.e., conformations of the bound complex) is used to approximate the time average of property A as shown in eq 3.[20]

$$\langle \mathbf{A} \rangle_{\text{ensemble}} = \langle \mathbf{A} \rangle_{\text{time}} \approx \frac{1}{M} \sum_{i=1}^{M} \mathbf{A}(t_i) \tag{3}$$

In our case, the property we are interested in is the potential energy of the system. The binding energy will be calculated as the difference between the average energy of the complex and the average energies of the ligand and receptor, as shown in eq 4.

$$\langle \Delta E \rangle_{\text{binding}} = \langle E \rangle_{\text{complex}} - \langle E \rangle_{\text{receptor}} - \langle E \rangle_{\text{ligand}} \tag{4}$$

For this reason, molecular dynamics simulation is necessary to generate a set of representative conformations of the ligand–protein complex, which can be used to assess the stability of the complex and more importantly to re-evaluate the scores of each of the eight binding modes obtained from the docking experiment.

2.2. Molecular Dynamics Simulations. Molecular dynamics simulations were performed for the eight ligand–protein complexes representing the eight different binding modes identified in the docking steps. The stability of the protein (tubulin), as well as that of the ligand (lankacidin C) in the binding site, was assessed by following the root-mean-
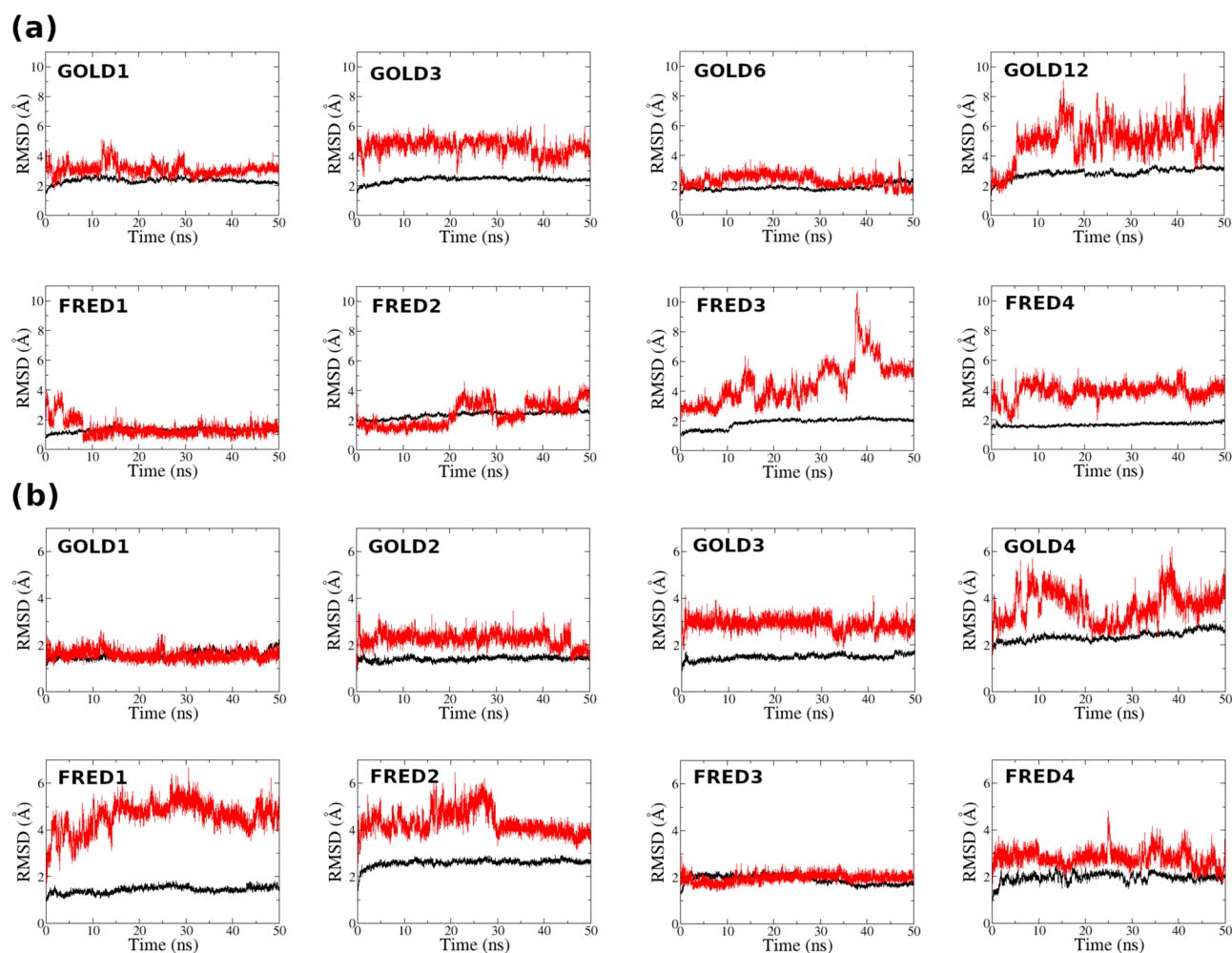
Figure 5. RMSD change over the course of the simulation for the different binding modes of (a) lankacidin C and (b) dictyostatin. The black lines represent the RMSD of the C_α atoms of the protein. The red lines (usually higher in the plots) represent the RMSD of the ligand atoms. All RMSD values were obtained after least-square fitting of the C_α atoms of the protein and with respect to the initial structure before the simulations.

squared deviation (RMSD) change over the course of the simulation time as shown in Figure 5a. The RMSD was measured after least-square fitting of the C_α atoms of the protein to that of the initial structure before the simulations. The figure shows that the RMSD of the protein (black lines) is stable for most of the complexes, ranging from 2.0 to 3.0 Å. The most stable protein among all of the complexes is the protein in the FRED1 binding mode where the RMSD of the protein is stabilized early in the simulation with very little fluctuations all around an average RMSD value of 1.3 Å. This means that for FRED1 binding mode, the protein stays for the whole 50 ns of the simulation in a conformation very close to the starting conformation obtained from 5LXT. Having an RMSD of less than 2.0 Å usually means that the conformation is almost unchanged.

Analyzing the change in RMSD of the ligand in Figure 5a, we find large values of RMSD as well as large fluctuations in the case of GOLD12, FRED3, FRED4, FRED2, GOLD3, and GOLD1 binding modes. The large values of RMSD (>2.0 Å) indicate that the binding mode of the ligand has changed over the course of the simulation time and it no longer resembles the initial binding mode. Not only are the values of RMSD large, but there are also large fluctuations in the values of RMSD over the course of the 50 ns of the simulation indicating that the ligand cannot attain a stable orientation in

the binding pocket. The ligand is stuck oscillating in the middle of some unfavorable orientations that cannot be easily changed to a favorable one due to the presence of high energy barriers. Therefore, it is unlikely that any of the six aforementioned modes will be the correct binding mode. For the two remaining modes, namely, FRED1 and GOLD6, the case is different. The RMSD fluctuations are small and the values are all around 2.0 Å, indicating that the ligand is stable in this binding mode and that the initial binding mode before the simulation is nearly conserved. The FRED1 binding mode is even more stable. Despite some fluctuations in the first 8 ns, the ligand remains around an RMSD value of nearly 1.2 Å until the end of the simulation, in contrast to an average value of 2.4 Å for GOLD6. This means that during the first 8 ns, the ligand and the protein in the FRED1 binding mode were slightly rearranged in space to accommodate each other better. After that, the ligand remained in an essentially unchanged orientation, very similar to the starting orientation, for the remaining 42 ns of the simulation. This is another evidence increasing the likelihood that FRED1 is the correct binding mode of lankacidin C to tubulin, despite the GOLD6 binding mode being a strong competitor.

Analyzing the RMSD plots of dictyostatin−tubulin complexes shown in Figure 5b, we can similarly rule out GOLD4, FRED1, FRED2, and FRED4. They are very unlikely to be the
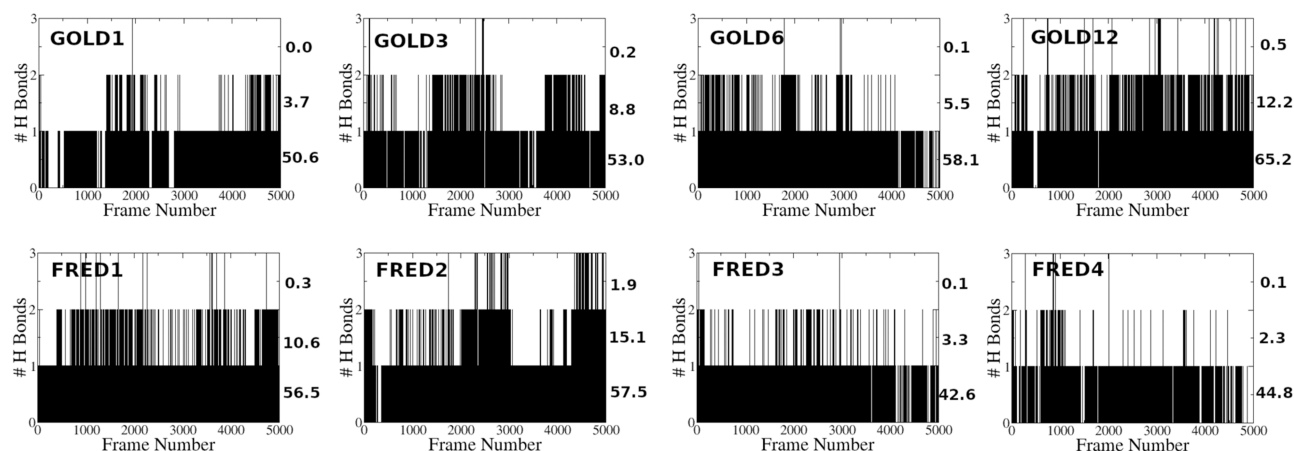
Figure 6. Results of the analysis of hydrogen bonds between lankacidin C and the protein along the 50 ns trajectories of the eight complexes. The *x* axis shows the frame number at a density of 100 frames/ns. The numbers to the right of each plot represent the persistence percentage of single, double, and triple hydrogen bonds (bottom to top) over the entire simulation time.

Table 2. MM/PBSA Binding Energies (BE) in kJ/mol between the Ligand (Lankacidin C or Dictyostatin) and the Protein in the Eight Binding Modes[a]

	lankacidin C					dictyostatin				
mode	vdW	ele	PS	NPS	BE	mode	vdW	ele	PS	NPS	BE
FRED1	−190	−41	152	−20	−99 ± 16	GOLD1	−213	−83	164	−22	−154 ± 17
FRED2	−203	−40	165	−21	−99 ± 17	GOLD2	−209	−52	157	−23	−127 ± 16
GOLD1	−133	−40	100	−17	−90 ± 13	FRED3	−242	−22	177	−25	−112 ± 16
GOLD6	−175	−36	145	−20	−86 ± 20	GOLD3	−173	−41	128	−20	−106 ± 18
GOLD12	−124	−43	104	−15	−78 ± 20	FRED2	−162	−25	106	−19	−100 ± 15
FRED3	−159	−25	125	−18	−77 ± 17	GOLD4	−189	−20	128	−22	−103 ± 21
GOLD3	−134	−56	132	−17	−75 ± 22	FRED4	−160	−25	107	−21	−99 ± 20
FRED4	−155	−39	156	−20	−58 ± 13	FRED1	−177	−11	113	−22	−97 ± 16

[a]The table shows different contributions to the binding energy including van der Waals (vdW) interactions, electrostatic interactions (ele), polar solvation (PS), and nonpolar solvation (NPS).

correct binding modes amid the instability reflected through their RMSD plots. To a lesser extent, GOLD3 and GOLD2 also are unlikely to be the correct binding modes since their ligand RMSD values are greater than 2.0 Å, and less but significant fluctuations can also be seen. We are left with GOLD1 and FRED3 binding modes, where the RMSD of both the ligand and the protein reflects a great deal of stability. In the case of GOLD1, the average value of ligand RMSD after the first 10 ns is 1.6 Å, versus 2.0 Å in case of FRED3. Indeed, the GOLD1 binding mode is the most stable among the eight proposed binding modes, which agrees with the fact that it is the correct binding mode. Once again, this strongly supports our approach in this study and highlights the reliability of using RMSD as a measure of binding mode stability and likelihood of being the correct one.

A hydrogen bond analysis was done to study the hydrogen bonds between lankacidin C and the protein along the 50 ns simulation trajectory of each of the eight complexes. The results are presented in Figure 6. The frequency of the hydrogen bonds over the simulation time as shown in the figure can indicate that the binding modes FRED3, FRED4, and GOLD1 are unlikely to be the correct binding modes due to the low frequency of hydrogen bonds. Other modes such as GOLD12, FRED2, FRED1, GOLD6, and GOLD3 have higher hydrogen bond frequencies, and one of them is more likely to be the correct binding mode. It is worth mentioning that the frequency of hydrogen bonds alone, without considering the

strength of each bond, is not enough measure for the strength of binding since few strong hydrogen bonds can make a bigger contribution to binding than many weak hydrogen bonds. Therefore, the results of this analysis can only be taken as a rough estimation. A deeper analysis of the binding energy can be achieved with molecular mechanics/Poisson−Boltzmann surface area (MM/PBSA) calculations.

2.3. MM/PBSA Binding-Energy Calculations. The binding energy was estimated using the MM/PBSA method, which takes into account contributions due to van der Waals (vdW) and electrostatic interactions (ele) (utilizing molecular mechanics, MM), polar solvation (PS) energy (utilizing Poisson−Boltzmann (PB) method), and nonpolar solvent-accessible surface area energy (utilizing a surface area term, SA). This, however, does not take into account the contributions due to rotational, translational, or vibrational entropy. Since in our study we are comparing the same ligand bound to the same protein but only in different binding modes, the entropic contributions are not expected to be much different between different binding modes. This is why the costly entropic calculations via normal mode analysis were not necessary and we depended on MM/PBSA alone to compare the different binding modes.

The results of MM/PBSA and its different contributions in the eight different binding modes of lankacidin C are shown in Table 2. As expected, the list is topped by the FRED1 binding mode with an average binding energy of −99 kJ/mol over the

course of the trajectory. An equal binding energy was also reported for FRED2 binding mode. All of the remaining binding modes have significantly weaker binding than the first two with the third-ranking mode, GOLD1, being statistically lower than the first two in a confidence interval greater than 99%. Most of that difference in binding energy comes from van der Waals interactions and polar solvation energy. With regards to dictyostatin, Table 2 shows that the GOLD1 binding mode is at the top of the list with an average binding energy of −154 kJ/mol. Thus, the binding in the case of the GOLD1 binding mode is significantly stronger than the binding in the case of all of the other binding modes. This is in line with all of the previous findings that GOLD1 is the most stable binding mode besides being the native binding mode found in the experimental structure. In turn, it is another evidence boosting the reliability of the protocol used in this study. In addition, upon comparing the binding energy of the best-scoring binding mode of dictyostatin (−154 kJ/mol) and the best-scoring binding mode of lankacidin C (−99 kJ/mol), we find that the results agree very well with experimental data. The IC$_{50}$ of the cytotoxic activity of dictyostatin lies in the nanomolar range,[19] whereas that of lankacidin C lies in the micromolar range.[14] Both our estimates of the binding energies and the experimental values of IC$_{50}$ reflect that dictyostatin has a cytotoxicity that is orders of magnitude stronger than that of lankacidin C.

2.4. Multiple-Trajectory Approach. To evade the risks that emerge from relying on a single molecular dynamics simulation, we ran a multiple-trajectory analysis to further assess the stability of the binding modes. In the multiple-trajectory simulations, five different parallel molecular dynamics runs were carried out for each of the eight binding modes of lankacidin C, each of the runs was 10 ns long. Each of the five parallel runs starts with a different velocity distribution, and hence the trajectories should be uncorrelated. Thus, we rule out the possibility that a single trajectory may be stable simply because of a specific favorable initial velocity distribution, or vice versa. As outlined in the methods section, a certain binding mode would be judged as a stable binding mode if at least one of the five parallel runs achieved an average ligand RMSD value <2.0 Å relative to the starting structure before the simulation. This criterion, as well as the entire method, was tested and validated recently.[21,22] The average RMSD values for all of the runs are listed in Table 3. As the table shows, the only binding mode that attained an average

RMSD value of <2.0 Å at least once is FRED1. In fact, not only one but two out of the five parallel simulations, namely, MD1 and MD2, satisfied this stability criterion. Therefore, it is the most likely binding mode compared to the others. All other binding modes did not satisfy this stability criterion, which indicates the very low likelihood that any of them could be the correct binding mode. Table 3, however, also shows that the MD3, MD4, and MD5 runs of the FRED1 binding mode did not satisfy our stability criterion. This behavior is acceptable due to the fact that each simulation run possibly explores a different region of the molecular dynamics phase space trajectory. Therefore, one should not expect every simulation run to achieve perfect stability, which does not happen even in real life as the ligand dynamically associates and then dissociates from the receptor. This is why the parallel molecular dynamics approach is preferable for such investigations and that is why our criteria for stability only required a minimum of one of the five simulations to achieve stability, as similarly argued by Liu and Kokubo.[21,22] Nevertheless, we wanted to investigate what could happen if we extend the five simulations of FRED1 to about 40 ns and measure the average RMSD in the last 10 ns of the simulation. We also did the same to the GOLD6 binding mode, which was the strongest competitor to FRED1 based on the previous RMSD analysis (see Figure 5). The results of the extended simulations indicated that for FRED1, the average RMSD values for both MD1 and MD2 remained below 2.0 Å, with values of 1.7 and 1.1 Å, respectively. MD5 also attained an average RMSD value of 1.0 Å early after the first 15 ns of the simulation. Therefore, three of the five extended runs meet our stability criterion. The average RMSD value of the MD3 run considerably declined from 5.6 Å to nearly 2.2 Å, but this does not meet our stability criterion. On the other hand, the average RMSD value of MD4 remained around 5.4 Å and never went down below 2.0 Å. For the GOLD6 binding mode, the average RMSD values of all of the five extended simulation runs remained above the cutoff of 2.0 Å, the lowest of which was the MD5 run, which had an average RMSD value slightly above 2.0 Å. The other four runs had average RMSD values ranging from 2.5 to 7.4 Å. Therefore, three out of the five extended runs for FRED1 showed adequate stability, whereas none of the five runs for GOLD6 met the preset stability criterion.

Given those findings, the FRED1 binding mode emerges as the most likely binding mode of lankacidin C in the taxol binding site. This mode showed a high score in docking, demonstrated a great deal of stability in RMSD analysis, performed decently in hydrogen bond analysis, and most importantly achieved the highest score in the MM/PBSA calculations as well as the multiple-trajectory approach. Despite the high score of the FRED2 binding mode of lankacidin C in the MM/PBSA analysis, its relatively poor performance in the RMSD analysis and the multiple-trajectory approach makes it less likely to be the correct binding mode due to the considerable instability of the ligand in its binding pocket. In addition, although GOLD6 performed well in RMSD analysis, its poor score in the MM/PBSA analysis and the multiple-trajectory approach makes it unlikely to be the correct binding mode. It is interesting to mention that the FRED1 binding mode of lankacidin C identified in this study is significantly different from the preliminary binding mode that was previously proposed[14,15] but never tested thoroughly, and this is part of the reason why this study was conducted.

Table 3. Average Lankacidin C RMSD Values of the Last Nanosecond in Each of the Five Parallel Molecular Dynamics Simulation Runs, Each of Which Is 10 ns Long[a]

mode	MD1	MD2	MD3	MD4	MD5
GOLD1	2.6	4.4	7.7	5.7	6.1
GOLD3	6.3	3.9	4.2	3.5	9.0
GOLD6	3.5	2.4	3.9	3.6	2.0
GOLD12	4.5	2.2	4.8	5.2	5.0
FRED1	**1.1**[b]	**1.0**[b]	5.6	5.6	2.3
FRED2	4.1	4.4	3.9	3.7	2.8
FRED3	2.0	6.5	5.0	4.5	2.9
FRED4	2.6	10.9	4.6	4.1	3.9

[a]Only the FRED1 binding mode can be judged as stable according to our definition of stability. [b]Bold entries highlight the RMSD values that are less than 2.0 Å and thus meet the stability criteria.
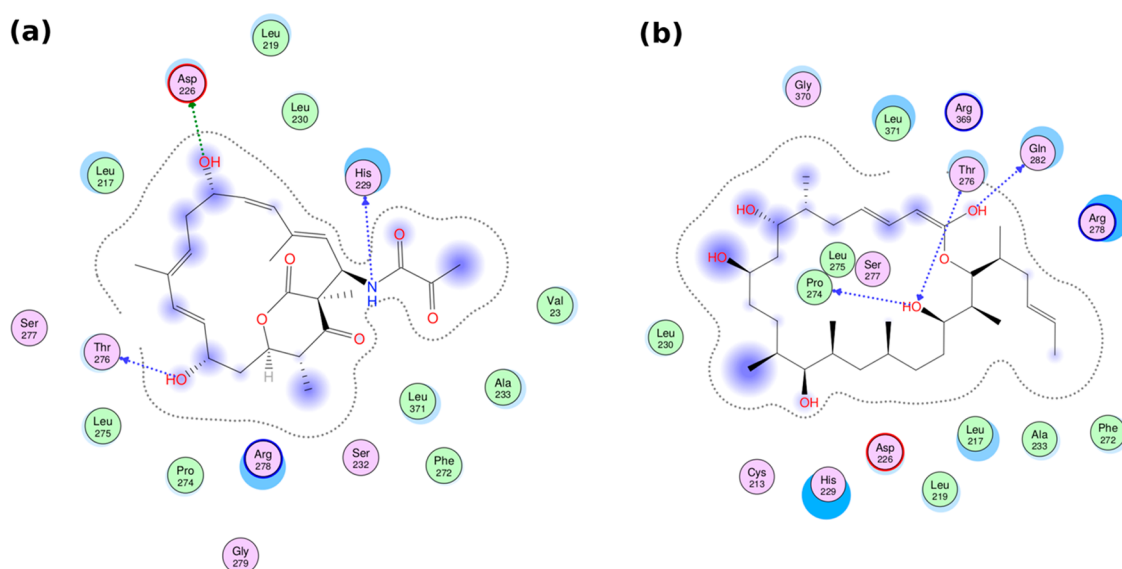
Figure 7. Interactions between the tubulin taxol binding site and (a) lankacidin C in the FRED1 binding mode or (b) dictyostatin in the native binding mode.

Visual analysis of the FRED1 binding mode, see the Supporting Information for an animation of the entire trajectory, reveals a ligand that is very stable in its binding pocket, very well buried inside the pocket with strong persistent interactions with nearby residues as was shown in the hydrogen bond analysis. Figure 7 shows the binding interactions of lankacidin C in the FRED1 binding mode as well as dictyostatin in the native binding mode inside the taxol binding site of tubulin. The figure shows that lankacidin C makes hydrogen bonds with the side chain of Asp226 and the backbones of Thr276 and His229, whereas dictyostatin makes hydrogen bonds with the backbones of Thr276, Gln282, and Pro274.

3. CONCLUSIONS

In this study, we have identified the most likely binding mode of lankacidin C in the taxol binding site of tubulin utilizing computational techniques. Eight different binding modes were proposed based on an ensemble-based docking experiment, which partially takes the receptor flexibility into account. After that, a molecular dynamics simulation was performed to assess the stability of each of the eight binding modes, allowing for full protein and ligand flexibility. RMSD analysis revealed that the FRED1 binding mode was the most stable among the binding modes. Hydrogen bond analysis showed that the FRED1 binding mode shows up on top of the list of hydrogen bond frequencies, despite being competed with GOLD12 and FRED2, both of which have already been ruled out due to instability in RMSD analysis. MM/PBSA binding-energy calculations revealed that the FRED1 binding mode is the best-scoring mode in terms of binding energy. Multiple-trajectory approach revealed that only the FRED1 binding mode met the required criterion for stability. Overall, the results indicate that the FRED1 binding mode is the most likely one among the eight studied binding modes. The coordinates of the FRED1 complex are available in the Supporting Information. The inclusion of dictyostatin in the study served as a benchmark for validating the reliability of the method for binding mode prediction. This study, however, needs to be confirmed through experimental techniques such

as X-ray crystallography, which is part of our future plans. Upon establishing the binding mode of lankacidin C to the taxol binding site, structure-based design techniques can immediately be applied to optimize the binding of lankacidin C to tubulin and improve its antitumor activity to compete with other antitumor drugs available in the market.

4. EXPERIMENTAL SECTION

This study was performed in four consecutive steps. First, an ensemble-based docking was performed to identify a group of potential binding modes of lankacidin C in the taxol binding site of tubulin. Second, the tubulin−lankacidin C complex of each of the different binding modes was studied using all-atom, explicit solvent molecular dynamics simulations. The molecular dynamics trajectories were studied to assess the stability of the complexes. Third, the trajectories were used to estimate the binding energy between the ligand and the protein in each binding mode using the MM/PBSA method. Finally, a multiple-trajectory approach was followed to further assess the stability of each binding mode. In addition, validating this protocol required that we apply the same steps to dictyostatin, whose native crystal structure with tubulin is available, to make sure that the protocol is able to predict its binding mode correctly. The reason we choose dictyostatin is its structural similarity to lankacidin C.

4.1. Ensemble-Based Docking. In ensemble-based docking, an ensemble of different receptor binding site conformations is used for docking, in an effort to account for receptor flexibility in the docking process. The ligand is docked into the binding site of each of these different conformations, and the different binding poses are ranked according to some scoring function. In this study and to obtain an ensemble of different tubulin binding site conformations, we screened the Protein Data Bank for entries of tubulin bound to various structurally diverse molecules at the taxol binding site. The search returned four distinct tubulin−ligand complexes including tubulin−paclitaxel complex with the PDB ID: 3J6G,[16] tubulin−discodermolide complex with the PDB ID: 5LXT,[17] tubulin−epothilone A complex with the PDB ID: 1TVK,[18] and tubulin−dictyostatin complex with the PDB ID:

5MFA.[19] The binding pocket of each of these complexes takes a different shape to accommodate the ligand; see Figure 2. Each of these complexes was cleaned by the elimination of the unnecessary α-tubulin subunit and retention of the β subunit where the binding pocket is located. Molecular Operating Environment (MOE, 2012)[23] was used for receptor preparation and the addition of missing hydrogen atoms via protonate3D, which optimizes the tautomeric forms of residues such as histidine and allows flipping of amide groups of side chains to optimize protonation.

The ensemble-based docking was done twice: once utilizing GOLD 5.6.3 (CCDC)[24] and another time utilizing FRED from the OEDocking 3.2.0 suite (OpenEye).[25,26] In each case, a slightly different protocol was applied to make sure we try as many different protocols as possible. In GOLD docking, the processed β subunits of the proteins from the four different PDB structures were superimposed and were ready for docking. As to the ligand, the crystal structure of lankacidin C was downloaded from the Protein Data Bank, PDB ID: 3JQ4.[7] MOE was utilized to perform a stochastic conformational search for lankacidin C to capture the flexibility of the 17-membered macrocyclic ring. A database of the low-energy conformations of lankacidin C was then used for docking against the ensemble of protein conformations superimposed by GOLD in the previous step. In the docking protocol, the binding site was defined to be within 6.0 Å of the co-crystallized ligands in the taxol binding site; pyramidal N and ring corners were allowed to flip; and CHEMPLP was used as a scoring function, without early termination to generate diverse solutions. The solutions were then analyzed visually, and the best-ranked poses were classified into a set of four distinct modes. The complex of each mode together with its corresponding tubulin conformation was later taken into a molecular dynamics simulation for further assessment of the complex stability and binding energy.

As to the docking step performed by FRED, the β subunit of the four different tubulin conformations was processed similarly using the make_receptor module. Lankacidin C was then subjected to a conformational search using OMEGA 3.2,[27] allowing the macrocycle to be flexible to explore different ring conformations. The ensemble of lankacidin C conformations was then docked separately to each of the four different tubulin receptor conformations. The different poses were scored using the Chemgauss4 scoring function to identify the best-scoring poses in each of the four docking runs. Out of each of these runs, we separated the best-scoring pose of lankacidin C in the complex with the respective tubulin receptor conformation, ending up with four potential binding modes. These binding modes were also taken into molecular dynamics simulations to assess their stability and estimate the binding energy. Thus, in total, we ended up with four binding modes from GOLD and another four binding modes from FRED. Similarly, we applied the same aforementioned docking protocol to dock dictyostatin, whose experimental structure is available, to the taxol binding site to assess the ability of the protocol to find the correct binding mode. We made sure that we randomize the structure of dictyostatin before docking to eliminate any bias in the docking process.

4.2. Molecular Dynamics Simulations. The docking step generated a set of potential binding modes for lankacidin C with different conformations of tubulin. A complex of the ligand and the corresponding tubulin conformation for each of the identified distinct binding modes was further analyzed by molecular dynamics. The GROMACS 2016.3 molecular dynamics package was used for this purpose.[28] We used the force field AMBER99SB-ILDN for the protein,[29] the GAFF force field for lankacidin C,[30] and the polyphosphate parameters developed by Meagher et al.[31] for parameterizing the guanosine diphosphate cofactor in β-tubulin. We utilized ACPYPE script to facilitate ligand parameterization.[32] Each complex was then solvated in a dodecahedral water box extending at least 10.0 Å in every direction, using the TIP3P water model. The system was then neutralized by the addition of sodium ions. Extra sodium chloride was added to adjust the salt concentration to 150 mM. The system was then minimized using the steepest descent algorithm. This was followed with 100 ps NVT equilibration at 300 K with position restraints on the protein and ligands. Subsequently, 100 ps of NPT equilibration at 1.0 bar was performed under the same position restraints. In all of the dynamics, a short-range cutoff of 12.0 Å was used for nonbonded interactions. In the end, the position restraints were released and a production phase of 50 ns followed under the same conditions as described before, saving coordinates every 10 ps. Thus, considering the eight different binding modes, we have performed a total of $8 \times 50 = 400$ ns simulations of the lankacidin−tubulin complex. All of these simulations were performed on NVIDIA Tesla K80 Graphical Processing Units on Bibliotheca Alexandrina High-Performance Computing Cluster (BA-HPC), Alexandria, Egypt. The trajectories were later analyzed by assessing the RMSD of the protein and ligand. This was done after least-square fitting of the C_α atoms of the protein and with respect to the initial structure before the simulation. Also, a hydrogen bond analysis between the ligand and the protein in each complex was performed over the entire trajectory, utilizing VMD 1.9.4a12,[33] using default values of a cutoff distance of 3.0 Å and a cutoff angle of 20°. Binding energies were analyzed as will be described in the following section. Similarly, this protocol was also applied to dictyostatin as a proof of concept.

4.3. MM/PBSA Binding-Energy Calculations. The molecular dynamics trajectory of each of the studied complexes was used to estimate the binding energy between the ligand and the receptor using the MM/PBSA method. This method divides the binding energy into three components assessed with three different methods. The first component is nonbonded interactions between the ligand and the receptor (van der Waals and electrostatic interactions), which is assessed using molecular mechanics (MM). The second component is the electrostatic contribution to the solvation free energy, which is assessed using the linearized Poisson−Boltzmann (PB) method. The third component is the hydrophobic contribution to the solvation free energy, which is calculated using a surface area (SA) term. For this sake, we have discarded the first 5 ns from each trajectory and used the remaining 45 ns for the MM/PBSA calculation at 250 ps intervals, for a total of 180 frames. We used the g_mmpbsa script developed by Kumari et al.[34,35] for the MM/PBSA calculations. We used a solute dielectric constant of 2 and a solvent dielectric constant of 80 for electrostatics. We used the default values for surface tension ($\gamma = 0.0226778$ kJ/(mol Å²)), probe radius (1.4 Å), and an offset constant of 3.84982 kJ/mol for the surface area term. Default values were used for the remaining parameters of the calculation. The binding energies of all different binding modes were compared, and the best-scoring mode was obtained. These calculations were done for both lankacidin C and dictyostatin.

4.4. Multiple-Trajectory Approach. A multiple-trajectory analysis was performed to further assess the stability of each binding mode of lankacidin C. All of the eight binding modes were taken into a molecular dynamics simulation using the same conditions and parameters described in the previous simulation. This time, for each binding mode, five independent parallel molecular dynamics simulation runs were performed, each having a production phase of 10 ns length. Thus, the initial velocity distributions generated in each of the five runs should be different and the trajectories should be uncorrelated. Hence, in total, $10 \times 5 \times 8 = 400$ ns simulations were performed in this step. After that, the trajectory of each run was analyzed by considering 100 snapshots from the last 1 ns of the 10 ns trajectory. The average RMSD value of the ligand heavy atoms was calculated over these 100 snapshots, after fitting the protein backbone atoms to the initial protein structure before the simulation. For each of the eight binding modes, the average RMSD value for each of the five independent simulations was assessed, and if the value of at least one of these five runs was less than 2.0 Å, the ligand was judged to be stable in this binding mode. This approach was outlined by Liu and Kokubo in previous studies and was shown to be effective in assessing the stability of different binding modes.[21,22] Based on the results of this analysis, some of the simulation runs were extended from 10 to 40 ns to further confirm persistent stability or instability. This resulted in a total of extra 300 ns simulation time. Collectively, the length of the molecular dynamics simulations applied to lankacidin C–tubulin complex in the parallel runs outlined in this section as well as the single runs outlined in Section 4.2 was 400 + 300 + 400 = 1100 ns or 1.1 μs.

ASSOCIATED CONTENT

ⓈSupporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b03470.

Visual animation of the molecular dynamics trajectory of the most likely binding mode, FRED1, showing lankacidin C in the binding pocket of tubulin (AVI)

PDB file showing the ligand–protein complex of the most likely binding mode identified in the study, FRED1 (PDB)

AUTHOR INFORMATION

Corresponding Author
*E-mail: atayoub@ualberta.ca.

ORCID ⓞ
Ahmed Taha Ayoub: 0000-0001-9106-1076

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Harada, S.; Higashide, E.; Fugono, T.; Kishi, T. Isolation and structures of T-2636 antibiotics. *Tetrahedron Lett.* **1969**, *10*, 2239–2244.

(2) Harada, S.; Kishi, T.; Mizuno, K. Studies on T-2636 antibiotics. II. Isolation and chemical properties of T-2636 antibiotics. *J. Antibiot.* **1971**, *24*, 13–22.

(3) Harada, S.; Tanayama, S.; Kishi, T. Studies on lankacidin-group (T-2636) antibiotics. 8. Metabolism of lankacidin C 14-propionate in rats and mice. *J. Antibiot.* **1973**, *26*, 658–668.

(4) Harada, S.; Yamazaki, T.; Hatano, K.; Tsuchiya, K.; Kishi, T. Studies on lankacidin-group (T-2636) antibiotics. VII. Structure-activity relationships of lankacidin-group antibiotics. *J. Antibiot.* **1973**, *26*, 647–657.

(5) Harada, S.; Kishi, T. Studies on lankacidin-group (T-2636) antibiotics. V. Chemical structures of lankacidin-group antibiotics. 1. *Chem. Pharm. Bull.* **1974**, *22*, 99–108.

(6) Tsuchiya, K.; Yamazaki, T.; Takeuchi, Y.; Oishi, T. Studies on T-2636 antibiotics. IV. In vitro and in vivo antibacterial activity of T-2636 antibiotics. *J. Antibiot.* **1971**, *24*, 29–41.

(7) Auerbach, T.; Mermershtain, I.; Davidovich, C.; Bashan, A.; Belousoff, M.; Wekselman, I.; Zimmerman, E.; Xiong, L.; Klepacki, D.; Arakawa, K.; Kinashi, H.; Mankin, A. S.; Yonath, A. The structure of ribosome-lankacidin complex reveals ribosomal sites for synergistic antibiotics. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 1983–1988.

(8) Belousoff, M. J.; Shapira, T.; Bashan, A.; Zimmerman, E.; Rozenberg, H.; Arakawa, K.; Kinashi, H.; Yonath, A. Crystal structure of the synergistic antibiotic pair, lankamycin and lankacidin, in complex with the large ribosomal subunit. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 2717–2722.

(9) Kende, A. S.; Liu, K.; Kaldor, I.; Dorey, G.; Koch, K. Total synthesis of the macrolide antitumor antibiotic lankacidin C. *J. Am. Chem. Soc.* **1995**, *117*, 8258–8270.

(10) Arakawa, K.; Sugino, F.; Kodama, K.; Ishii, T.; Kinashi, H. Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in *Streptomyces rochei*. *Chem. Biol.* **2005**, *12*, 249–256.

(11) Zheng, K.; Shen, D.; Hong, R. Biomimetic synthesis of lankacidin antibiotics. *J. Am. Chem. Soc.* **2017**, *139*, 12939–12942.

(12) Ootsu, K.; Matsumoto, T. Effects of lankacidin group (T2636) antibiotics on the tumor growth and immune response against sheep erythrocytes in mice. *Gann* **1973**, *64*, 481–492.

(13) Ootsu, K.; Matsumoto, T.; Harada, S.; Kishi, T. Antitumor and immunosuppressive activities of lankacidin-group antibiotics: structure-activity relationships. *Cancer Chemother. Rep.* **1975**, *59*, 919–928.

(14) Ayoub, A. T.; Abou El-Magd, R. M.; Xiao, J.; Lewis, C. W.; Tilli, T. M.; Arakawa, K.; Nindita, Y.; Chan, G.; Sun, L.; Glover, M.; Klobukowski, M.; Tuszynski, J. Antitumor activity of lankacidin group antibiotics is due to microtubule stabilization via a paclitaxel-like mechanism. *J. Med. Chem.* **2016**, *59*, 9532–9540.

(15) Ayoub, A. T.; Klobukowski, M.; Tuszynski, J. Similarity-based virtual screening for microtubule stabilizers reveals novel antimitotic scaffold. *J. Mol. Graph. Model.* **2013**, *44*, 188–196.

(16) Alushin, G. M.; Lander, G. C.; Kellogg, E. H.; Zhang, R.; Baker, D.; Nogales, E. High-resolution microtubule structures reveal the structural transitions in $\alpha\beta$-tubulin upon GTP hydrolysis. *Cell* **2014**, *157*, 1117–1129.

(17) Prota, A. E.; Bargsten, K.; Redondo-Horcajo, M.; Smith, A. B.; Yang, C.-P. H.; McDaid, H. M.; Paterson, I.; Horwitz, S. B.; Fernando Díaz, J.; Steinmetz, M. O. Structural basis of microtubule stabilization by discodermolide. *ChemBioChem* **2017**, *18*, 905–909.

(18) Nettles, J. H.; Li, H.; Cornett, B.; Krahn, J. M.; Snyder, J. P.; Downing, K. H. The binding mode of epothilone A on α,β-yubulin by electron crystallography. *Science* **2004**, *305*, 866–869.

(19) Trigili, C.; Barasoain, I.; Sánchez-Murcia, P. A.; Bargsten, K.; Redondo-Horcajo, M.; Nogales, A.; Gardner, N. M.; Meyer, A.; Naylor, G. J.; Gómez-Rubio, E.; Gago, F.; Steinmetz, M. O.; Paterson, I.; Prota, A. E.; Díaz, J. F. Structural determinants of the dictyostatin chemotype for tubulin binding affinity and antitumor activity Against

taxane- and epothilone-resistant cancer cells. *ACS Omega* **2016**, *1*, 1192−1204.

(20) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons Ltd.: Canada, 2004.

(21) Liu, K.; Watanabe, E.; Kokubo, H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. *J. Comput.-Aid Mol. Des.* **2017**, *31*, 201−211.

(22) Liu, K.; Kokubo, H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations: A cross-docking study. *J. Chem. Inf. Model.* **2017**, *57*, 2514−2522.

(23) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: Montreal, QC, Canada, 2012.

(24) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(25) McGann, M. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578−596.

(26) McGann, M. FRED and HYBRID docking performance on standardized datasets. *J. Comput.-Aid Mol. Des.* **2012**, *26*, 897−906.

(27) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.

(28) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to super-computers. *SoftwareX* **2015**, *1−2*, 19−25.

(29) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950−1958.

(30) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(31) Meagher, K. L.; Redman, L. T.; Carlson, H. A. Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem.* **2003**, *24*, 1016−1025.

(32) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **2012**, *5*, 367.

(33) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.

(34) Kumari, R.; Kumar, R.; Lynn, A. g_mmpbsa-A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **2014**, *54*, 1951−1962.

(35) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.