PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Flexibility and inhibitor binding in Cdc25 phosphatases

Guilherme Menegon Arantes*

Instituto de Química, Universidade de São Paulo, Av. Lineu Prestes 748, 05508-900, São Paulo, SP, Brasil

ABSTRACT

Cdc25 phosphatases involved in cell cycle checkpoints are now active targets for the development of anti-cancer therapies. Rational drug design would certainly benefit from detailed structural information for Cdc25s. However, only apo- or sulfate-bound crystal structures of the Cdc25 catalytic domain have been described so far. Together with previously available crystalographic data, results from molecular dynamics simulations, bioinformatic analysis, and computer-generated conformational ensembles shown here indicate that the last 30–40 residues in the C-terminus of Cdc25B are partially unfolded or disordered in solution. The effect of C-terminal flexibility upon binding of two potent small molecule inhibitors to Cdc25B is then analyzed by using three structural models with variable levels of flexibility, including an equilibrium distributed ensemble of Cdc25B backbone conformations. The three Cdc25B structural models are used in combination with flexible docking, clustering, and calculation of binding free energies by the linear interaction energy approximation to construct and validate Cdc25B-inhibitor complexes. Two binding sites are identified on top and beside the Cdc25B active site. The diversity of interaction modes found increases with receptor flexibility. Backbone flexibility allows the formation of transient cavities or compact hydrophobic units on the surface of the stable, folded protein core that are unexposed or unavailable for ligand binding in rigid and densely packed crystal structures. The present results may help to speculate on the mechanisms of small molecule complexation to partially unfolded or locally disordered proteins.

Proteins 2010; 78:3017–3032.

Key words: molecular dynamics; conformational ensemble; docking; intrinsic disorder; protein tyrosine phosphatases; transient cavities; computer simulation; receptor flexibility.

INTRODUCTION

Cell division cycle Cdc25 phosphatases have turned into active targets for the development of antineoplastic agents as their overexpression has been detected in several human cancers.[1–7] Although different classes of Cdc25s inhibitors have been identified so far, the number of potent inhibitors that are active *in vivo* is still limited and restricted to quinoid-based compounds.[1,2] Thus, rational drug design toward Cdc25 would certainly benefit from a detailed knowledge of this enzyme catalytic mechanism and tertiary structure.

Cdc25s are dual-specificity protein tyrosine phosphatases (PTP) that dephosphorylate pTyr and pThr residues on cyclin-dependent kinases (Cdk) responsible for cell cycle checkpoints.[8,9] Known as the P-loop, the active site is characterized by the sequence Cys-$(Xxx)_5$-Arg conserved among all known PTPs (Fig. 1). The conserved Cys is the nucleophile that attacks the substrate phosphate group, which is in turn stabilized by the conserved Arg side chain. Poor selectivity has been observed *in vitro* for the most potent PTP inhibitors because these compounds bind to the conserved P-loop.[4,5] One important difference between Cdc25 and the other PTPs that could be explored in the design of selective inhibitors is the catalytic general acid. In Cdc25, this residue was identified by the Glu following the nucleophilic Cys in the P-loop sequence.[9,10] But, in all other PTPs, the general acid is an Asp located in a flexible loop distant from the active site.[8,11]

The three Cdc25 isoforms found in humans share 70% identity in the primary sequence of their catalytic C-terminal domain, but present larger variance in their N-terminal regulatory domain where several splice variants are possible.[1] All the structural information available for Cdc25s was obtained from X-ray crystallographic data for Cdc25A[12] and Cdc25B[13] catalytic domains in either apo forms (PDB code 1c25)[12] or with sulfate bound to the P-loop (codes 1qb0, 1cwr, 2a2k, 2ifd).[13,14] NMR structures and order parameters[15] for Cdc25 in solution, structures for inhibitor-Cdc25 complexes, or for the divergent N-terminal domain have not been described to our knowledge. The most notable difference between the crystal structures of isoforms Cdc25A and B is observed in the backbone of the last 20 aminoacids in the C-terminus. This region contains an α-helix for Cdc25B but it is unstructured for Cdc25A (Fig. 1). Additional 28 (Cdc25A) and 16

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: FAPESP; Grant numbers: 07/52772-6, 07/59345-6
*Correspondence to: Guilherme Menegon Arantes, Instituto de Química, Universidade de São Paulo, Av. Lineu Prestes 748, 05508-900, São Paulo, SP, Brasil. E-mail: garantes@iq.usp.br
Received 4 April 2010; Revised 7 June 2010; Accepted 23 June 2010
Published online 23 July 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.22826

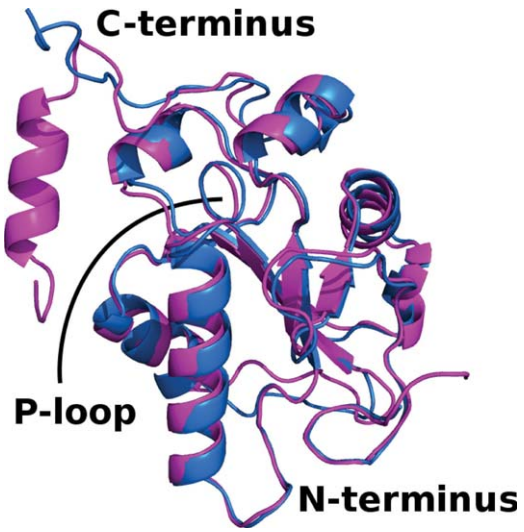

Figure 1

Superposition of the Cdc25A (in blue, PDB code 1c25) and Cdc25B (in magenta, PDB code 1qb0) catalytic domain crystal structures.

(Cdc25B) residues in the protein C-terminal were present in the constructs used for crystallization, but had no observable electron densities suggesting further intrinsic flexibility or disorder in the C-terminus.[12,13] Interestingly, it has been shown that this C-terminal tail mediates substrate recognition by up to 100-fold in the phosphatase activity of Cdc25B and Cdc25C toward the bisphosphorylated Cdk2-pTpY/CycA natural substrate.[16]

Structural disorder plays a role in the molecular recognition between proteins involved in cell signaling.[17] In particular, the interaction between Cdk and several of its partner proteins including p21,[18] but not yet Cdc25, has been shown to proceed through disorder-to-order transitions.[19] Thus, it is possible that the recognition between Cdc25 and Cdk/Cyc complexes is driven by a similar mechanism involving the last 30–40 residues in the Cdc25 C-terminal. Even if a disorder-to-order transition is not involved, C-terminal flexibility will be relevant because this tail occupies the protein-protein contact surface between the Cdc25 catalytic domain and the Cdk2 main body.[14] The Cdc25 C-terminus is also in contact with its own main body and near the active site P-loop (Fig. 1), such that its flexibility will also influence binding of competitive inhibitors. What is then, the effect of this putative C-terminal disorder on the complexation of small molecules to Cdc25? It appears that this question has received little attention for partially unfolded proteins in general, even though it has been suggested that a large fraction of eukariotic proteins (~30%) contains intrinsically disordered regions.[17]

Modelling protein–inhibitor complexation, in particular, docking of small molecules to rigid receptor structures, is an established area with several parameterized and tested methods.[20,21] Including protein flexibility into such docking protocols is, however, less developed.[22,23] Most of the approaches currently in use were designed to account for side chain or limited backbone flexibility, for instance, by introducing soft or average grid models of receptor ligand interaction.[24,25] This is clearly inappropriate for large-scale fluctuations present in partially disordered proteins because average positions will not represent the relevant conformational states.[23,26] Receptor flexibility may also be included by docking to several different structures from a configurational ensemble.[27–29] The problem here is how to obtain representative structures for flexible or disordered protein domains. A set of geometries obtained from X-ray crystallography might be perturbed by crystal packing[30,31] and disordered segments will not have visible electron densities. Ensembles obtained from NMR might be more appropriate models for solution configurations but the fitting procedure traditionally used for structural determination by reproducing distance restrains becomes less reliable for proteins with intrinsic disorder and does not result in an equilibrium distribution of configurations.[15,30]

Receptor structures may be generated computationally, but large-scale fluctuations and protein disorder are more difficult to simulate accurately. Molecular dynamics (MD) simulations up to the nanosecond timescale under normal temperature probe local fluctuations around the starting geometry.[32] High temperature MD may unveil cryptic binding sites or induced fit changes if the protein region that undergoes conformational exchange is known in advance.[33] Otherwise, several MD acceleration[34,35] and generalized ensemble[36,37] methods may be used to increase configurational sampling. Computer generation of protein conformational ensembles may be based on geometrical information of reference configurations,[38–40] on backbone dihedral angle propensities extracted from existing libraries,[41,42] or on combinations of both approaches.[43] Statistical libraries built from geometries of coiled protein fragments are reasonable models for the unfolded or disordered regions.[41,44] For protein–inhibitor complexation a further challenge is the selection of an ensemble subset that is relevant for binding.[22] Although a criteria based on structural variance is often used,[22,32] a more formal filtering should rely on the Boltzmann distribution of the ensemble.[45]

In this work, Cdc25B C-terminal flexibility is analyzed, and its effect on the complexation of quinone-based small molecules is investigated using computer simulations. Details about the computational methods used are given in the next section. Results for molecular dynamics simulations of Cdc25B strongly support that the C-terminus is flexible and partially unfolded in solution. Three Cdc25B structural models with variable levels of flexibility are compared: A rigid model corresponding to the crystal structure, a semiflexible model corresponding to a set of MD snapshots and a fully flexible model built with

statistical rigor for backbone configurations. Then, Cdc25B-inhibitor complexes using the three structural models are presented and validated by the calculation of binding free energies. Given the lack of experimental data to benchmark the structural models and the generated ligand–receptor complexes, the focus here is on discussing the binding sites and less detail is reported on specific ligand–receptor molecular interactions. Finally, the effect of C-terminal flexibility on the small molecule complexation is described.

METHODS

Molecular dynamics and system set-up

Coordinates of a sulfate-bound Cdc25B crystal structure deposited in the PDB code 1qb0 were used to start the molecular dynamics simulation. This protein chain contains 177 residues from Asp374 to Trp550 of the Cdc25B primary sequence. The initial simulation model was built by removing all crystallographic waters, ions, and β-mercaptoethanol. Missing hydrogen atoms were added, considering all Asp, Glu, Arg, Lys, and His side chains charged. This protonation state is slightly different from the one accepted as the catalytic competent state in which Cys473 is deprotonated and Glu474 protonated.[9] Such difference should not influence the flexibility in the P-loop active site backbone, which, is rather rigid, but it might play a role for the detailed interactions in small molecule complexation in the active site. The protein chain was solvated by superimposing it to a replicated water box, leaving at least 12 Å between the protein and the box edge. Six chloride anions were added to neutralize the total system charge. Initial coordinates for one of the chloride anions were extracted from the crystal structure (PDB code 1qb0). Although it was suggested that this chloride anion may have a structural role,[13] it diffuses out of the interaction site within 5 ns of simulation. Periodic boundary conditions were used with a cubic box length of ~78 Å. Long-range electrostatics were treated with the particle-mesh Ewald summation with a 1.2 Å Fourier grid spacing and fourth-order interpolation. A switch function with r_{off} = 10.0 Å was used to truncate long-range van der Waals interactions without dispersion corrections.

All model building, molecular dynamics simulations and trajectory analysis were carried out with the GROMACS 4.0.2 suite of programs.[46] The OPLS-AA[47] force field was used for the protein and ions, and the rigid TIP3P[48] potential was used for water. The LINCS algorithm was used to constrain all covalent bonds and a time step of 2 fs was used in MD integration.

The solvated protein heavy atoms were restrained to their initial coordinates. The system was first energy-minimized to relax unfavorable contacts, and kinetic energy was gradually injected in the system by running short (40 ps) MD simulations at 50, 100, 200, and 300 K. Restrains in the protein heavy atoms were gradually removed by reducing the harmonic constant from the initial 1000 kj mol^{-1} $Å^{-2}$ to 200, 20, and 2 kj mol^{-1} $Å^{-2}$ along another set of short (40 ps) MD simulations at 300 K. This careful heating and pre-equilibration phase was used to avoid artifactual flexibility in the apo Cdc25B simulation as well as artifactual decomplexation of inhibitors in the evaluation of interaction energies (see Linear Interaction Energy section below). An equilibration 2 ns MD run was obtained without any position restrains at 300 K and 1 atm. The Berendsen temperature and pressure couplings were used up to this point. Finally, a production 60 ns MD trajectory was obtained at the same temperature and pressure conditions but in a canonical ensemble using the Nose-Hoover extended temperature scheme (time constant τ_T = 0.1 ps) and the Parrinello–Rahman pressure coupling (time constant τ_P = 0.5 ps). No group separation was used.

Bioinformatic analysis of disordered sequences

Four different bioinformatic methods were used to predict the disordered regions of Cdc25 isoforms A and B. Primary sequences were submitted to web servers queries and default algorithm options were used. FoldIndex[49] builds a score based on residue charge and hydrophobicity for a sliding window of 51 residues. Disopred2[50] uses a linear support vector machine to analyze the sequence profiles. VL3H[51] and VLXT[52] both search using neural networks, and the latter is based on residue physicochemical properties. Default server and algorithmic values were used in all queries. Only VL3H gives a continuous N-terminal disordered segment. All other predictors give N-terminal segments containing ordered gaps of up to 30 amino acids, but with a total of more than 75% of disordered residues in the N-terminal domain. The score given for the discontinuous segments is the average of disordered scores (FoldIndex and VLXT) or the range of scores (Disopred2). In FoldIndex, negative scores mean a segment will more likely be disordered. Scores on the other methods give the probability of disordered structure in the segment.

Ensemble generation

Two conformational ensembles based on the Cdc25B catalytic domain crystal structure (code 1qb0) were used here. The MD ensemble was composed of 150 snapshots spaced by 400 ps each from the 60 ns MD trajectory described above. Another Cdc25B conformational ensemble built with a library-based Monte Carlo approach was kindly provided by D. Zuckerman and co-workers.[43] The 177 Cdc25B residues deposited on the PDB code 1qb0 are present in this model. The protein is divided into

fragments identified with residue types for which statistical libraries of backbone conformations were built in advance. Fragment libraries have Boltzmann distributions p_i^{frag}, which account for correlations internal to each fragment. The distribution of protein configurations built with a simple product of fragments

$$p_i^{\text{lib}}(\mathbf{r}_i, \ldots, \mathbf{r}_M) = \prod_i^M p_i^{\text{frag}}(\mathbf{r}) \quad (1)$$

will be biased toward the isolated fragment distributions. Thus, Metropolis Monte Carlo is used to resample and obtain protein configurations in equilibrium distribution.[53] The protein internal energy is written as:

$$U^{\text{prot}}(\mathbf{r}_i, \ldots, \mathbf{r}_M) = \sum_i^M U_i^{\text{frag}}(\mathbf{r}_i) + U^{\text{rest}}(\mathbf{r}_i, \ldots, \mathbf{r}_M) \quad (2)$$

where U_i^{frag} is the energy internal to each fragment obtained from the library and U^{rest} accounts for the interactions between the M fragments. The interaction term is described by a structure-based (or Gō type) potential, which stabilizes the native Cdc25B fold[13] but allows fluctuations around it in reasonable agreement with experimental data.[43] It should be noted that Boltzmann distributed protein backbone configurations are generated with this formal procedure. However, residue side chain conformations are built for each backbone geometry separately, using the backbone-dependent rotamer library SCWRL.[54] Thus, side chains will not follow the equilibrium distribution. Further details for the generation of the Zuckerman Cdc25B ensemble are given in the original paper.[43]

Small molecule docking and clustering

The studied ligands were 6-chloro-7-(2-morpholin-4-ylethylamino)quinoline-5,8-dione (NSC 663284), **1** in Figure 2, and 2,5-dihydroxy-3-(7–(2-methyl-benzyl)-1H-indol-3-yl)[1,4]benzoquinone, **2** in Figure 2. For each receptor structure, 200 AutoDock 4.0 runs were performed for each ligand.[55] Interaction maps were generated in a cubic grid with 60 points and 0.375 Å spacing, centered in the Cys473 S_γ, the catalytic nucleophile. This medium sized grid allowed docking to sites up to 11 Å away from the Cdc25B active site. A distance-dependent dielectric constant was used for electrostatic interactions.[55] Four torsions for **1** and five torsions for **2** were active. The genetic algorithm used for conformational search was set with 150 individuals in the population for a maximum of 27,000 generations or 2.5×10^6 energy evaluations.

Two clustering steps were used to analyze the docking poses (more than 30.000 per ligand per ensemble). Within each receptor structure, poses were initially clustered for ligand heavy atom root mean-squared deviation (RMSD) with a 2.0 Å cutoff. The pose with lowest free energy score in each initial cluster was considered a valid one if it belonged to a cluster with at least 20 members

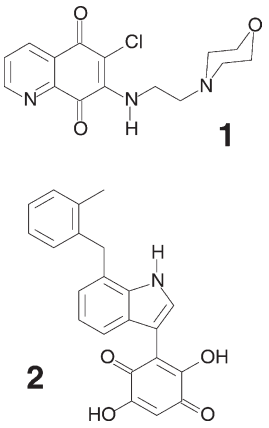

Figure 2
Structure of the quinolinequinone **1** and the indolyldihydroxyquinone **2**.

and if its AutoDock 4.0 free energy score was at most 2.0 kcal/mol higher than the lowest free energy score found within dockings for the same receptor structure. This is the definition for valid pose or valid complex used throughout this work. Then, the set of valid poses for all receptor structures in each ensemble was clustered for ligand heavy-atom RMSD tolerance of 4.0 Å after least-square fitting of C_αs for the stable residues in the receptor (residues 377–530). This procedure gave the best compromise between the size of clustering problem, the amount of clusters finally obtained, and the expected accuracy of the docking function.[55] The eight clusters with the largest population of valid poses for each ligand and ensemble were analyzed in detail below.

Several criteria in the above procedure were relaxed to test for cluster robustness. For instance, decreasing the number of members (20) or increasing the cutoff energy (2 kcal/mol) in 50% for the first clustering step, enlarged the set of valid poses, but yielded qualitatively the same eight most populated clusters, that is, the same binding sites and cluster centroids. Throughout this work, the centroid geometry represents the clusters obtained in the second clustering step. A cluster centroid is defined as the valid complex with the smallest average distance to the other valid complexes within the same cluster. Again, the same binding sites were obtained if the RMSD tolerance was changed to 3.5 or 4.5 Å in the second clustering step (the population of each cluster was smaller or larger, respectively). Several clustering algorithms (single-linkage, Daura *et al.*,[56] and self-organizing maps[57]) were tested yielding similar results.

Linear interaction energy

The stability of the binding modes found in the docking and clustering procedure was estimated using the fol-

lowing linear interaction energy (LIE) approximation to the binding free energy[58]:

$$\Delta G_{\text{bind}} \simeq \alpha(\langle V_{\text{comp}}^{\text{vdW}} \rangle - \langle V_{\text{sol}}^{\text{vdW}} \rangle) + \beta(\langle V_{\text{comp}}^{\text{elet}} \rangle - \langle V_{\text{sol}}^{\text{elet}} \rangle) \quad (3)$$

where $\langle \cdots \rangle$ represents an ensemble average (obtained here from a MD trajectory) of nonbonded van der Waals (vdw) and electrostatic (elec) interaction energies (V) between the ligand and its surrounding environment. Direct calculation of absolute free energies is possible,[59] but extremely demanding for computational resources and, hence, unfeasible for such large set of complexes as found here. The LIE is computationally more efficient because it only requires two end states MD simulations, one for the free ligand in solvent (sol) and another for the ligand–protein complex (*comp*). The parameters adopted were $\alpha = 0.18$, $\beta = 0.37$ for **1**, and $\beta = 0.33$ for **2**.[60] MD simulations starting from the ligand–protein complex poses obtained with AutoDock and for the free ligand in water were carried out following the same model building, heating, and pre-equilibration procedures adopted for the apo Cdc25B (Molecular dynamics and system set-up section). The equilibration and data acquisition MD trajectories were run for 0.5 and 1.0 ns, respectively, and both simulations were obtained at 300 K and 1 atm in a canonical ensemble using the Nose-Hoover and the Parrinello–Rahman coupling schemes. Long-range electrostatics were treated with a switch function with $r_{on} = 8$ Å and $r_{off} = 12$ Å. The force field for both ligands was derived from the OPLS-AA force field with appropriate partial charges derived from the charges assigned to similar OPLS groups.

RESULTS AND DISCUSSION

Cdc25B is partially unfolded in solution

To explore the intrinsic flexibility of apo Cdc25B in aqueous solvent, a long molecular dynamics simulation was conducted at normal conditions. The observed trajectory begins with a rapid relaxation of the protein structure from the initial crystallographic coordinates in response to the solvated environment. The protein core (residues 374–531) has a constant deviation from the initial coordinates measured by a C_α RMSD ~2 Å up to the end of the simulation as shown in Figure 3(A). However, at about 40 ns, a steep increase in the C_α RMSD from 4 Å to more than 8 Å is observed for the last 19 residues in the Cdc25B C-terminus [Fig. 3(A)]. The protein radius of gyration, which is stable at ~16.7 Å from 5 to 40 ns also increases to ~17.0 Å from 40 ns to the end of the simulation.

Figure 3(B) shows fluctuations of the protein C_αs about their mean position obtained from the Cdc25B MD run and from the temperature factors of Cdc25A and Cdc25B crystallographic structures. The Cdc25B

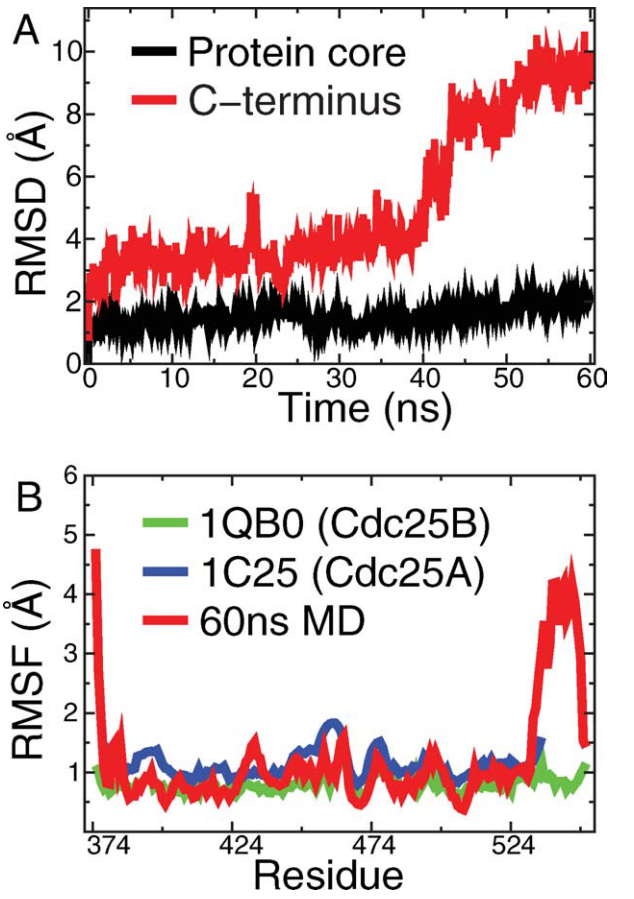

Figure 3

Molecular dynamics of apo Cdc25B in solution. Top panel A shows the C_α root mean-squared deviation from the crystalographic structure for the Cdc25B protein core (residues 374–531) and the 19 residues in the C-terminus (residues 532–550). Lower panel B gives the C_α root mean-squared fluctuation obtained from the 60ns MD and calculated from temperature factors for the Cdc25A (PDB code 1c25) and Cdc25B (PDB code 1qb0) structures. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

RMSF obtained from the X-ray structure is rather flat and featureless, which maybe a consequence of the refinement procedure on the diffraction data.[31] There is some fair agreement between the RMSF obtained from the Cdc25A structure and from the Cdc25B simulation. The catalytic domains of the two isoforms share more than 70% identity in their primary sequences and have almost identical folds (Fig. 1), so they should also have equivalent backbone fluctuations. For instance, both RMSFs are larger than 1 Å from residues 443–466, and drop steadily in residues 469–471. This is the active site region, which is prearranged to bind phosphate esters.[9,11,13] The main differences between fluctuations calculated from the experimental and simulated data are observed for the protein terminals. The simulated RMSF is >3 Å for the first three residues (374–376) in the N-

terminus and for residues 531–548 in the C-terminus [Fig. 3(B)]. The large and localized flexibility in the N-terminus should be expected as the protein is truncated in this position. In fact, the Cdc25B construct expressed and used for X-ray crystallography has in addition 18 residues (356–373) along the N-terminus and 16 residues (551–566) along the C-terminus, which are not observed in the electron density maps.[13] The simulated protein contains only residues 374–550, as given in the PDB file, including an α-helix present along residues 533–546.[13]

The large RMSD and RMSF observed for the C-terminus during the MD trajectory correspond to a local unfolding and detaching of the terminal α-helix from the protein main-body. It may be possible that the last 16 residues (551–566) along the C-terminus, which were present in the Cdc25B construct used for crystallization, but absent from the MD simulation, are necessary to stabilize the terminal α-helix. However, no terminal helix is observed in the crystal structure for isoform Cdc25A (Fig. 1) even though the complete C-terminus, residues 336–523 in the Cdc25A numbering, was present in the crystallized construct. In fact, electron densities for the last 28 residues (496–523) in Cdc25A were not observed and this region was considered disordered in the crystal form.[12]

These observations suggest that the terminal Cdc25 regions, especially the last 30–40 residues in the C-terminus, are intrinsically disordered or at least partially unfolded in solution. The terminal α-helix present in the Cdc25B structure may be a crystal packing artifact or may be stabilized because favorable interactions between symmetrically related molecules (the terminal Trp550 side chains are in contact in the 1qb0 model[13,43]).

To support the partial unfolding hypothesis, the primary sequences of Cdc25A and Cdc25B were submitted to bioinformatic analysis for prediction of disordered protein sequences.[52] Table I shows the results obtained with four different methods. The whole N-terminal domain is predicted as disordered, with up to 320 unfolded residues for Cdc25A and 370 residues for Cdc25B. To our knowledge, the structure of the N-terminal region of Cdc25s has never been solved, which might be explained by the difficulty in obtaining crystals and a diffraction pattern from partially disordered materials. The ordered segment found by all bioinformatic methods roughly coincides with the catalytic domains of Cdc25A and B observed by X-ray crystallography[12,13] and with the core region of Cdc25B stable during the MD simulations described above. The disordered segments found for the C-terminus are all continuous, but with variable size. They range from the last 75 residues (FoldIndex) down to the last 40 (VL3H) or 30–20 residues (Disopred2 and VLXT) for both Cdc25A and B. But, clearly, the bioinformatic search confirms the suggestion from MD simulations that the Cdc25 terminals are disordered or partially unfolded in solution.

Table I
Disordered Regions on the Cdc25A and B Complete Primary Sequences Predicted by Various Bioinformatic Methods[a]

	Cdc25A		Cdc25B	
Method	Segment	Score	Segment	Score
Disopred2[50]	1–323	0.2–1.0	1–368	0.2–1.0
	504–523	0.5	546–566	∼0.5
FoldIndex[49]	61–325	−0.20 ± 0.09	160–375	−0.22 ± 0.08
	450–523	−0.19 ± 0.08	494–566	−0.14 ± 0.06
VL3H[51]	1–323	0.8	1–368	0.8
	480–523	0.8	528–566	0.7
VLXT[52]	1–329	0.7	1–365	0.8
	493–523	0.8	541–562	0.9

[a]Residue numbering covers the entire protein sequence: 1–523 for Cdc25A and 1–566 for Cdc25B.[13]

Modeling Cdc25B with conformational ensembles

The effects of protein flexibility upon binding and molecular recognition may be accounted for by modeling the receptor geometry with a conformational ensemble or a collection of structures.[22,29] Statistically significant sampling of such conformations following a Boltzmann distribution[45] is not feasible with the computer power available nowadays if proteins the size of Cdc25 catalytic domain (∼200 residues) are modelled by detailed all-atom force fields, such as the one used here for MD. Different methods exist to generate conformational ensembles.[38–42,61,62] Here two computational models were adopted: An ensemble selected from the present MD trajectory and an ensemble generated for the Cdc25B catalytic domain by Zuckerman and co-workers.[43] See the Methods section for details about ensemble construction. The Zuckerman ensemble contained 250 structures with C_α RMSD from the X-ray structure varying from 3 to 10 Å. The core of the Cdc25B catalytic domain (residues 377–530) remains stable and folded. But, large fluctuations are again observed in the C-terminus. In fact, the C-terminal helix is unfolded in all configurations in the Zuckerman ensemble. This is another independent computational evidence of the intrinsic disorder suggested above for the Cdc25 C-terminal region. Figures shown below for ligand–protein complexes with the Zuckerman ensemble are illustrative examples of configurations composing this ensemble. The average radius of gyration for protein conformations in the ensemble is 17.8 ± 0.7 Å, ∼1 Å larger than the radius of gyration observed in the MD simulation. Even though the C-terminus unfolds, the main distance from residues to the protein center-of-mass does not change significantly, because the protein core remains stable and folded. In agreement with experimental evaluation of the Cdc25B crystal structure, part of the Zuckerman ensemble had a disulfide bridge formed between C426 and C473 (corresponding to the PDB model 1cwr).[13] Only the 176 structures without such bridge were used here.

Figure 4(A) shows the projections of the Zuckerman ensemble and of 1200 snapshots from the 60 ns MD trajectory spaced by 50 ps (including the MD ensemble) on the first and second eigenvalues of the C_α covariance matrix of the join set of configurations. These two principal components (PC) describe collective motions mainly by the last 20 residues (531–550) modeled in the Cdc25B C-terminus. The configurational space spanned by the Zuckerman ensemble (~14 nm for PC1 and PC2) is considerably larger than the space visited by the MD trajectory (~3 nm for PC1 and PC2). The MD trajectory overlaps the same four to eight ensemble structures in the first six PCs calculated [only PC1 and PC2 are shown in Fig. 4(A)]. The cosine content of PC1 is lower than 20% and it is lower than 1% for the other initial PCs.[63] These results suggest that the Zuckerman ensemble correctly sampled relevant solution conformations (the MD structures) but also sampled a much larger configurational space corresponding to large scale fluctuations of the disordered C-terminus. Figure 4(B) shows a contact map representation obtained from averaged distances for both ensembles. The tertiary structure is very similar between residues 374 and 530 for both ensembles. The largest differences are observed for contacts involving the C-terminus. The terminal helix is partially formed in the MD ensemble and in contact with the protein main body, particularly to residues 428–431, 480–483, and 445–446. In the Zuckerman ensemble, the C-terminus makes little to no contacts with the protein main body.

Deviations were calculated for both ensembles in comparison with the Cdc25B crystallographic structure. For the backbone atoms in the stable region (residues 377–530), the average RMSD is 1.1 ± 0.1 Å and 2.9 ± 0.1 Å for the MD and Zuckerman, respectively. The difference between the two ensembles is not large and similar to the deviations often observed between different crystallographic structures and solution NMR ensembles of the same protein.[30] For C_α of residues 531–550 in the C-terminus, two RMSDs distributions are clearly observed for the MD ensemble, one for configurations visited up to 40 ns, with average RMSD = 3.5 ± 0.7 Å, and another after 40 ns, with average RMSD = 8.2 ± 1.3 Å. The C-terminus average RMSD is 18.3 ± 6.0 Å for the Zuckerman ensemble. Such large RMSDs are a consequence of C-terminus local unfolding as detailed above. For the side chains of 29 residues in the surface of the stable protein region that may interact with small-molecule ligands (see next section), the average RMSD is 3.0 ± 0.4 Å and 3.7 ± 0.1 Å for the MD and Zuckerman, respectively. Together with Figure 4, these RMSD numbers indicate the degree of conformational variability between the two ensembles and the crystal structure.

Given the lack of structural information for Cdc25 in solution or bound to small-molecule inhibitors (see Introduction), which could be used to compare with the present ensembles and with the protein–ligand complexes generated by modelling, it is difficult to assess the quality of each Cdc25 structural representation. Thus, three models with variable levels of flexibility are identified and used in the next section for analysis of protein–ligand complexes: The rigid model corresponds to the crystal structure (1qb0) and contains a frozen and folded C-terminus; A semiflexible model corresponds to the MD ensemble and represents a distribution between a folded and a partially unfolded C-terminus; A fully flexible model with the C-terminus totally unfolded corresponds to the Zuckerman ensemble.

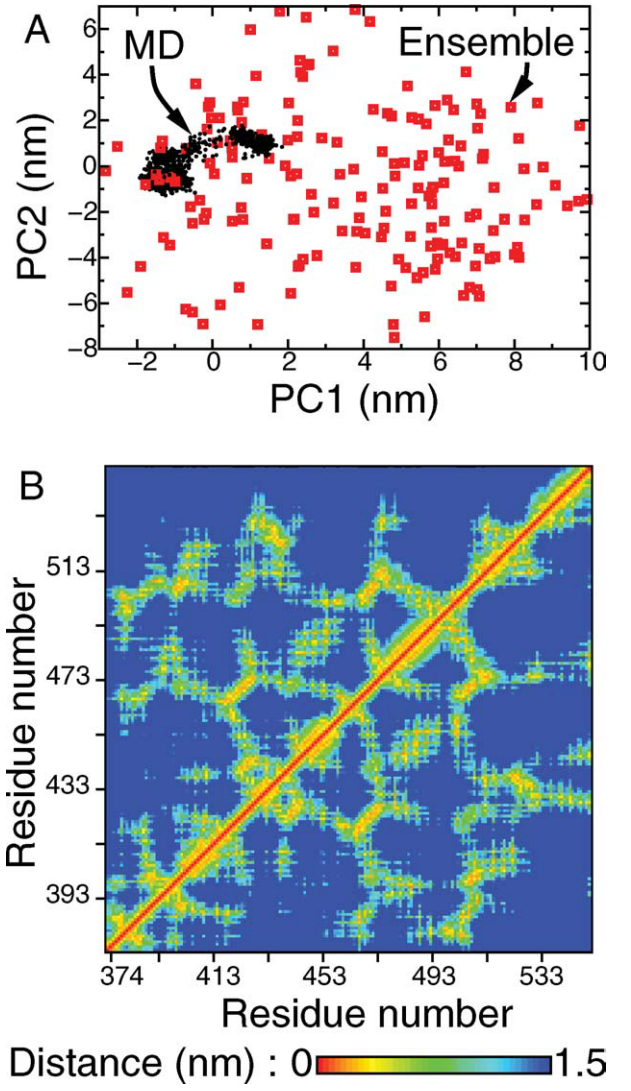

Figure 4

Comparison of the conformational ensembles used here. Top panel A shows the projections along the first and second eigenvalues obtained from principal component analysis of the MD trajectory (black) and the Zuckerman ensemble (red). Lower panel B shows a contact map representation of the MD ensemble (lower right half) and of the Zuckerman ensemble (upper left half) with smaller mean distances between residues color coded. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Each model has distinct structural characteristics and may suffer from different artifacts. The RMSF profiles obtained from experiment and from the MD simulation [Fig. 3(B)] indicate that the local fluctuations Cdc25B undergoes in the crystal are significantly different from fluctuations in solution suggesting that the crystallographic structure might not be a good model for the protein geometry in native conditions. Crystal packing and near-neighbor interactions in the crystal lattice may significantly alter the conformational distribution.[30,31] In extreme cases, the PDB model, which represents an average of this altered distribution, may not be an appropriate model for a flexible protein in solution. In addition, the rigid model will not be able to account for any induced fit caused by ligand complexation.

Although the MD ensemble was obtained from an MD trajectory that follows the equilibrium distribution, the ad hoc selection of a subset of this trajectory, either as structures equally spaced in time (as used here and often in ensembles derived from MD[32]) or in RMSD distance[22] may artificially bias the distribution. The MD ensemble will also suffer from undersampling for large-scale fluctuations in the Cdc25 C-terminus.[23] The semi-flexible model will, however, partially account for induced fit effects. Configurations induced by ligand binding should also be present in the apo state ensemble, although possibly in smaller populations.[64–66] Actually, by observing whether the population of ligand-Cdc25 complexes is enriched by structures obtained before or after 40 ns in the MD trajectory, it will be possible to tell if inhibitor complexation shifts the population distribution towards a state with folded or partially unfolded C-terminus, respectively.

In the Zuckerman ensemble, the large C-terminal fluctuations might be an artifact of the incomplete backbone sequence used in the ensemble generation or caused by an inappropriate choice of simulation parameters, in special the reduced temperature. The distribution of side chain dihedrals might also be incorrect, particularly for side chains exposed to solvent (and in contact with putative ligands).[67,68] The protocol used to build the side chain orientation for each backbone configuration is based on average rotameric information collected from apo crystal structures.[54] Nevertheless, assuming that the Cdc25 C-terminus is unfolded in solution as suggested in the previous section, the fully flexible representation will be the best model to analyze the effect of flexibility on small molecule complexation because it samples fluctuations of the C-terminus in a statistically rigorous fashion for the protein backbone.

Binding modes for complexation to Cdc25 models

Ligands **1** and **2** (Fig. 2), two of the most potent Cdc25B inhibitors known to date,[6,69,70] were docked to the Cdc25B PDB structure (code 1qb0) as well as to the MD and Zuckerman ensembles using the AutoDock empirical energy function.

Table II
Summary of the Valid Poses and Clusters Found for Docking to Each Cdc25B Structural Model

Model	1qb0	MD	Zuckerman
Ligand **1**			
Valid Poses	3	361	426
Poses per structure	3	2.4	2.4
Number of clusters		48	52
Poses in clusters 1–4		46%	45%
Poses in clusters 1–8		63%	65%
Ligand **2**			
Valid Poses	1	389	405
Poses per structure	1	2.6	2.3
Number of clusters		38	56
Poses in clusters 1–4		52%	48%
Poses in clusters 1–8		74%	67%

A large number of docking poses was found for each ensemble. Table II shows results of the clustering procedure used to aggregate and classify valid poses (defined in Small molecule docking and clustering section). The average number per ensemble structure indicates that two to three valid poses were obtained for each receptor geometry. Throughout the remainder of this article, each cluster of valid poses will be identified as a possible Cdc25B-ligand binding mode. Up to 50 different clusters were found for each ligand-ensemble pair, but most of these contain very few (1–5) valid poses. Half of the valid poses are aggregated in the four most populated clusters and two-thirds are in the eight most populated clusters.

Because of the lack of structural information about Cdc25-inhibitor complexes, a pragmatic criteria combining high population and favorable energy score was used to choose clusters for further analysis: The eight most populated clusters for each ligand-ensemble pair were validated and rescored based on an estimate of the binding free energy. Several studies suggest that the most populated clusters are better predictors of the native binding mode than the docking pose with most favorable score.[28,71] This may be viewed as an averaging procedure that compensates noisy scoring functions and is based on the assumption that the global free energy minimum contains the largest amount of structural neighbors, increasing the available configurational entropy.[72]

As the energy functions used for docking are rather crude,[23,55,73] a more accurate estimate of the binding free energy was necessary to rescore the clusters found. Absolute binding free energies could be calculated exactly for a given force field by statistical perturbation or thermodynamic integration.[59] However, the slow convergence and high computational cost preclude using these methods for the large set of complexes studied here. Instead, the linear interaction energy (LIE) estimate of

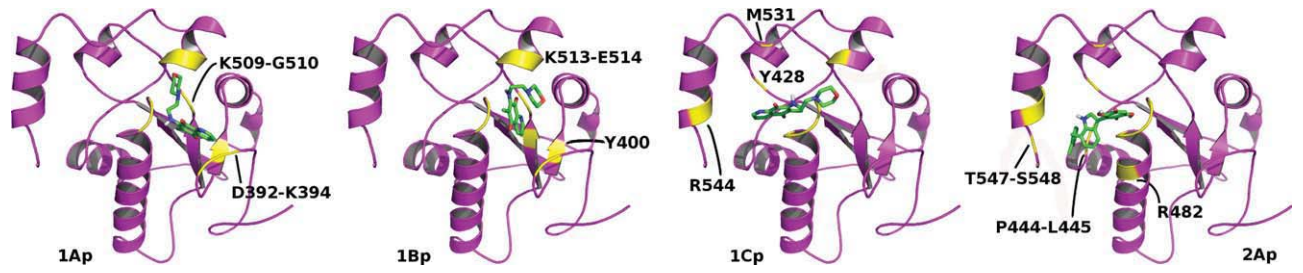

Figure 5

Complexes between the Cdc25B crystal structure and **1** (three structures on the left) and **2** (structure on the right). Residues in contact with the ligand (distance <4.5 Å) are shown in yellow and printed accordingly. P-loop residue numbering was not printed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the binding free energy was used because it is a more economic but still reliable free energy model.[58,74] LIE approximations have their roots on physical theories of solvation linear response and, in principle, can yield absolute free energies.[58] However, the formulations used in practice, such as eq. 3, resemble linear structure-activity relationships which are empirically trained by adjusting coefficients [(α and β in Eq. (3)]. Thus, parameterized LIE relationships are often used to rank a series of related ligands and its accuracy therefore depends on the similarity between ligands. Because **1** and **2** are structurally different and the LIE relationship used here was not trained for these specific ligands, we do not expect to discriminate the most potent between the two inhibitors, neither to obtain LIE estimates in close agreement with experimental binding free energies. However, the LIE values are well-suited to rank the various clusters or binding modes found within each ligand-structural model pair because errors will largely cancel between complexes with the same ligand.[58] Because the LIE includes both receptor and ligand local flexibility, a detailed all-atom force field description for the energy function and explicit solvent contributions, its ranking accuracy is expected to be superior to the AutoDock energy function.[58]

The number of valid poses used in the LIE evaluation for each cluster was composed by the maximum between 15% of the cluster size and 4 poses. The cluster centroid was included in the set, and the remaining poses were randomly chosen within each cluster.[73] For the rigid model, LIE scores represent averages $\pm$ standard deviations obtained for four geometries in the same valid pose. Complexes were considered relevant binding modes in the discussion below only if their average LIE score less one standard deviation was smaller (or more stable) than the lowest average LIE plus one standard deviation within the same ligand-structural model pair. Only binding modes with the lowest intrinsic binding free energies (or highest partial association constant) will contribute significantly to the measured inhibition constants.[64,75]

Figure 5 shows the three valid poses found for **1**, and one valid pose found for **2** by docking to the rigid crystallographic model. LIE scores for complexes 1Ap, 1Bp, and 1Cp are -12.4 ± 1.0, -6.3 ± 1.8 and -12.6 ± 2.2 kcal/mol, respectively. Complex 1Bp will not be relevant because its intrinsic binding free energy is 6.0 kcal/mol lower. In complex 1Ap, the ligand is placed in a shallow pocket formed beside the P-loop, with the quinolinedione moiety in contact with residues D392-K394, Y400 and K509, and the morpholine group in contact to K513-E514. In 1Cp, the inhibitor is docked to the P-loop, with the quinolinedione also in contact to residues Y428, M531, and R544. Complex 2Ap LIE score equals to -7.9 ± 1.4 kcal/mol and shows the dihydroxyquinone group complexed with the P-loop and M531, the indole ring in contact with Y428 and R482, and the 2-methylbenzyl group coordinated to the C-terminus.

AutoDock and LIE results for binding free energies differ qualitatively. For instance, AutoDock gives similar ranking for complexes 1Ap, 1Bp, and 1Cp, but the LIE value for 1Bp is much higher than for the other two complexes. Similar qualitative differences are observed between AutoDock and LIE scores for the ensemble complexes. Tables III and IV give average LIE and standard deviation, as well as the number of valid poses found for the eight most populated clusters in each ligand-ensemble pair. Standard deviations from the average LIE energy are appreciable. RMSDs for the ligand atom positions between initial and final geometries out of the LIE MD trajectories vary from 0.5 to 1.7 Å, without any correlation to LIE deviations. The LIE deviations may be due to a lack of convergence in the average potential energies used in Eq. (3) and obtained from the relatively short (1.5 ns) MD trajectories.[58,60] For complexes with the ensemble, LIE deviations are also caused by inherent variability in initial geometries among valid poses within each cluster.

Normalized histograms depicting contacts between ligand and protein were built for complexes with both ensembles (four figures in the Supporting Information). A contact was defined if at least one residue atom lays at a distance $\leq$4.5 Å from any ligand atom. A list of contacts was generated for each valid pose and the fraction of each contact found within a cluster was calculated.

Table III
Properties of the Eight Most Populated Clusters Found for **1** Complexed to Cdc25B Ensembles

MD ensemble								
Complex	1Am	1Bm	1Cm	1Dm	1Em	1Fm	1Gm	1Hm
Valid poses	91	31	24	19	17	15	15	14
Poses used for LIE	14	5	4	4	4	4	4	4
Average LIE[a]	−5.4	−2.8	−0.6	−9.3	−7.7	−4.8	−11.0	−5.0
LIE Std. Deviation[a]	2.7	2.6	1.2	1.9	0.6	2.5	1.8	1.2
Zuckerman ensemble								
Complex	1Az	1Bz	1Cz	1Dz	1Ez	1Fz	1Gz	1Hz
Valid poses	73	51	39	27	26	22	20	19
Poses used for LIE	11	8	6	4	4	4	4	4
Average LIE[a]	−6.7	−10.8	−5.6	−8.8	−5.8	−10.5	−7.6	−6.1
LIE Std. Deviation[a]	2.4	1.7	2.1	2.1	1.2	2.2	1.0	2.5

[a]Average binding free-energy and standard deviation from LIE calculation given in kcal/mol.

The histograms identify the main contacts formed and their dispersion within each cluster.

From the eight most populated clusters found for binding between **1** and the MD ensemble, only 1Dm and 1Gm are identified as relevant modes by LIE scoring. Geometries for the cluster centroids representing these complexes are shown in Figure 6. Although its LIE score is rather low, 1Am is also shown because it corresponds to the most populated cluster found (Tables III and IV). In 1Am, the ligand is complexed to the P-loop and with the C-terminus. In 1Dm, the quinolinedione is buried in the same shallow pocket as in 1Ap. In 1Gm, the quinolinedione is complexed in front of the P-loop and the morpholine is buried inside the P-loop. Notably, all complexes, but 1Dm and 1Em, show contacts between **1** and W550 (See Supporting Information).

Figure 7 shows the centroid geometry for 4 out of the 5 clusters identified as relevant binding modes for complexes between **1** and the Zuckerman ensemble. The 8 most populated clusters found can be roughly divided in two groups. In 1Az and 1Dz, the ligand assumes different orientations but binds to the shallow pocket formed beside the P-loop and also observed in complexes 1Ap and 1Dm. In fact, 1Az and 1Ap correspond to almost identical binding modes with equivalent ligand-residue contacts. The other group is formed by 1Bz, 1Fz, and 1Gz and shows the ligand bound into the P-loop. In 1Gz (geometry not shown), the ligand occupies the same site as in 1Bz and 1Fz, but with morpholine coordinated to the P-loop and the quinolinedione in contact with residues M531-H533.

Thus, the same two binding sites are observed for **1** complexed with Cdc25B in the three structural models analyzed. In complexes 1Ap, 1Dm, 1Az, and 1Dz, the shallow pocket formed beside the P-loop together with residues D392-H395, Y400, K509-G510, and K513-E514 is occupied. It may accommodate two orientations of the quinolinedione group: complexed with the P-loop (complex 1Dz) or with residues D392-H395 (1Ap, 1Dm, and 1Az). The morpholine group may be exposed to solvent (1Dm) or in contact with K513-E514 (1Ap and 1Az). Backbone flexibility in the C-terminus and the protein main-body has little influence on binding to the shallow pocket since all structural models can account for binding. This is an obvious conclusion given the site distance from the C-terminus and the stability of the protein core (see Cdc25B is partially unfolded in solution section). However, the precise orientation and the contacts formed by **1** depend on the flexibility of P-loop and shallow pocket side chains.

The second binding site observed in 1Cp, 1Am, 1Gm, 1Bz, and 1Fz shows the quinolinedione coordinated to the P-loop in several orientations. The contacts formed and the ligand orientation are largely influenced by the C-terminus backbone flexibility. In the rigid model, dense packing by the C-terminus complexed with the protein main-body and side chains of Y428, M531, and R544 leaves little room for **1** so that the morpholine group points to the shallow pocket beside the P-loop and coordinates K513 (complex 1Cp). The increased receptor flexibility modeled by the MD ensemble corresponds to a partial detachment of the C-terminus from the protein main-body and allows ligand reorientation. Therefore, the morpholine group may coordinate deeper in the direction of residues E431 (1Gm) or Y528 (1Am). Although the C-terminus fluctuates widely in the fully-flexible model, the region composed of residues M531-N532 moves on average to configurations closer to the P-loop and makes favorable contacts with **1** (1Bz and 1Fz). Such fluctuations also allow placing the morpholine group in the space occupied by the C-terminus in the more rigid models. Backbone flexibility thus yields ligand complexation with regions of the stable protein core that were unavailable and unexposed in the rigid model.

Table IV
Properties of the Eight Most Populated Clusters Found for **2** Complexed to Cdc25B Ensembles

MD ensemble								
Complex	2Am	2Bm	2Cm	2Dm	2Em	2Fm	2Gm	2Hm
Valid poses	71	58	41	32	23	23	21	18
Poses used for LIE	11	9	6	5	4	4	4	4
Average LIE[a]	−8.9	−6.2	−8.7	−9.4	−11.5	−6.7	−6.8	−7.2
LIE Std. Deviation[a]	1.4	0.8	1.6	2.2	1.4	1.6	0.9	2.6
Zuckerman ensemble								
Complex	2Az	2Bz	2Cz	2Dz	2Ez	2Fz	2Gz	2Hz
Valid poses	67	60	37	32	28	18	15	13
Poses used for LIE	10	9	6	5	4	4	4	4
Average LIE[a]	−7.3	−8.9	−8.8	−10.0	−8.4	−7.7	−8.0	−9.8
LIE Std. Deviation[a]	1.3	0.8	1.6	2.5	2.3	1.9	1.4	1.3

[a]Average binding free-energy and standard deviation from LIE calculation given in kcal/mol.

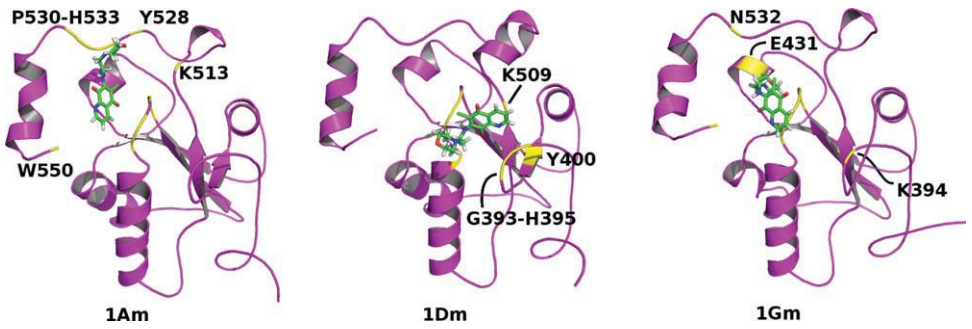

Figure 6

Relevant complexes found between **1** and the MD Cdc25B ensemble. Residues in contact with the ligand (distance <4.5 Å) and with a normalized count higher than 5% (see Fig. 11 and explanation on text) are shown in yellow and printed accordingly. P-loop residue numbering were not printed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Noteworthy, the LIE scores suggest that 1Cp, 1Gm, and 1Bz/1Fz are more stable than 1Ap, 1Dm, and 1Az/1Dz, respectively in each structural model. Thus, binding **1** to the P-loop should be more stable than binding to the shallow pocket.

Complexes between **2** and the MD ensemble that represent relevant binding modes are shown in Figure 8. In 2Am, the dihydroxyquinone is coordinated above the P-loop with the 2-methyl-benzyl moiety in contact with the C-terminus. In both 2Cm and 2Em, the ligand is complexed outside the P-loop, with the 2-methyl-benzyl group coordinated to the shallow pocket. Complex 2Em is the most stable and the only binding mode that does not coordinate to W550. In 2Dm, the dihydroxyquinone group coordinates both the P-loop and N532.

Complexes between **2** and the Zuckerman ensemble are shown in Figure 9. All of them have LIE scores within one standard deviation from the cluster with lowest average LIE. Therefore, the eight complexes may represent relevant binding modes. In both 2Az and 2Bz, **2** is complexed to the shallow pocket formed beside the P-loop. Complex 2Gz shows **2** coordinated with the C-terminus and only R479 in the P-loop. In the remaining complexes, the ligand binds to the P-loop. Similar contacts, excluding those with the C-terminus, are observed in 2Fz and 2Ap.

Because **2** is a bulkier ligand, it does not fit entirely in the shallow pocket identified above for **1**. Docking to the rigid crystal model results in only one binding mode where **2** is complexed to the P-loop and the C-terminus (2Ap). Increasing receptor flexibility leads to complexes in which (part of) **2** coordinates the shallow pocket (2Cm, 2Em, 2Az, and 2Bz). In fact, **2** fits completely in the shallow pocket only in the fully flexible model (2Az and 2Bz). Thus, backbone flexibility in the C-terminus grants access of bulkier groups to cavities transiently formed on the surface of the stable folded protein core.[66] Complexes coordinating the shallow pocket are less stable than complexes bound to the P-loop, as observed for **1**. Increased flexibility also results in a larger diversity of modes for **2** complexed with the P-loop.

For the semiflexible model, the fraction of receptor structures in each cluster obtained after 40 ns in the MD trajectory suggests whether complexation shifts the popu-

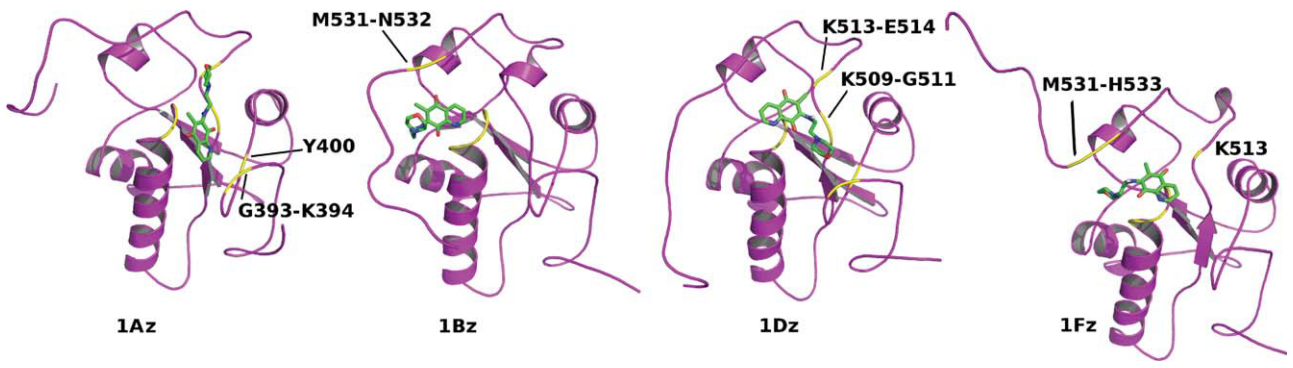

Figure 7

Relevant complexes found between **1** and the Zuckerman Cdc25B ensemble. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

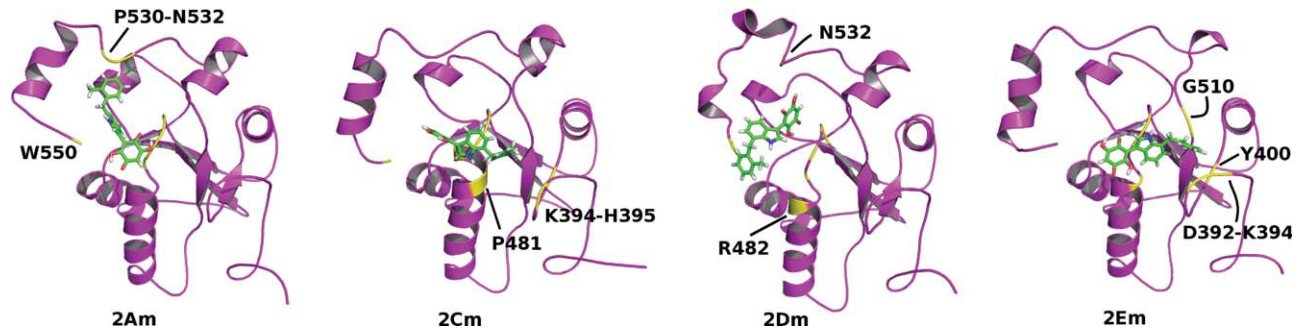

Figure 8

Relevant complexes found between **2** and the MD Cdc25B ensemble. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

lation distribution toward a state with a partially unfolded C-terminus.[65,66] Because one-third of the MD ensemble corresponds to structures obtained after 40 ns, a similar fraction would be expected in a cluster without a shifted receptor population. Modes 1Dm and 1Gm have fractions of 0.26 and 0.67, respectively. As expected from the complex geometries (Fig. 6), 1Dm does not have a clear dependence on the C-terminal flexibility. But, 1Gm shows significant enrichment for receptor structures corresponding to larger fluctuations of the C-terminus. Ligand coordination in 1Gm requires a receptor with less dense packing between the C-terminus and the protein main-body. For **2**, 2Dm is the only relevant complex with a fraction (0.59) significantly different from one-third, also in agreement with observation that the ligand requires more orientation freedom than available in the rigid model.

Comparison with previously available data

A series of quinolinediones were tested experimentally for Cdc25 inhibition.[69] The most potent inhibitor found was **1** (IC_{50} = 0.21 μM), 15-fold more potent than the unsubstituted quinolinedione. Several substitutions of the

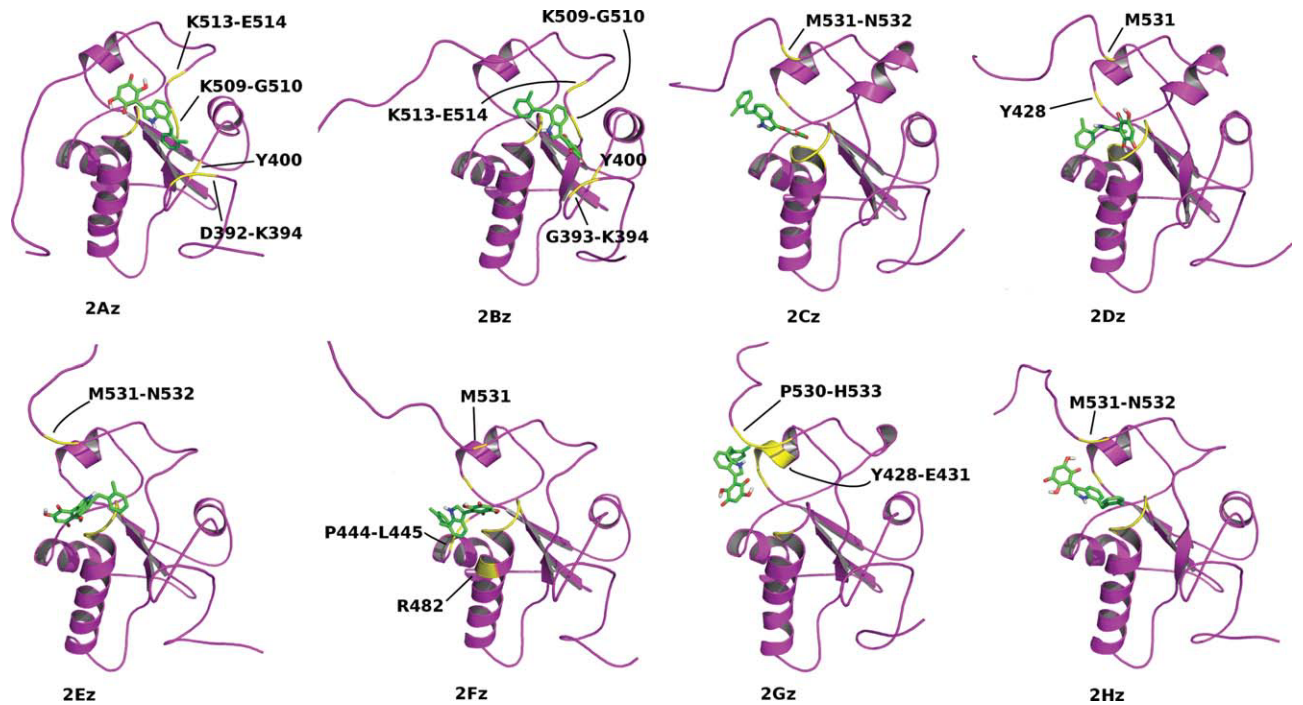

Figure 9

Relevant complexes found between **2** and the Zuckerman Cdc25B ensemble. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

morpholine-ethylamino group for other similar sized amino derivatives, removal or substitution of chloride by methoxide as well as different quinones (isoquinoline, quinazoline, etc.) resulted in 2- to 10-fold decrease in inhibition potency. The regioisomer of **1** with chloride and morpholine-ethylamino group positions exchanged was four-fold less potent. Thus, the quinolinedione is the main function responsible for inhibitory activity and the specific contacts of substituents have relatively less importance. Kinetic data for inhibition by **1** fit to a partial mixed competitive model[69] and displays a time-dependent inhibition.[70] As it was shown that **1** irreversibly oxidizes the catalytic C473 of Cdc25B,[76] it is plausible that **1** occupies two different binding sites in Cdc25: one for reversible, competitive binding and another for irreversible oxidation. Average edge-to-edge distances between the catalytic C473 and **1** are 3.3 $\pm$ 0.6 Å and 7 $\pm$ 1 Å for the binding mode on top of the P-loop (complexes 1Cp, 1Gm, 1Bz, and 1Fz) and in the shallow pocket beside the active site (1Ap, 1Dm, 1Az, and 1Dz), respectively. Thus, both sites may allow fast electron transfer for irreversible oxidation of C473.[77] On the other hand, only the complexes with **1** on top of the P-loop would display kinetics consistent with competitive inhibition.

Several indolyldihydroxyquinones related to **2** were screened for Cdc25 inhibition.[70] Modifications in the 2- and 4-position of the indole ring (position 1 corresponds to the nitrogen atom and 3 corresponds to the bond with quinone) lead to loss of inhibitory potency by hindering the torsion around the quinone bond. Modifications at 5- and 6-position have little effect, indicating that these positions do not make specific contacts with the receptor. Substitutions at 7-position with hydrophobic groups bulkier than propyl increase potency suggesting that the 2-methyl-benzyl group in **2** occupies a hydrophobic pocket. Two related inhibitors were screened against various Cdc25B point mutants: Molecule **2A** (in the nomenclature used in the original article[70]) lacks the 2-methyl-benzyl group at 7-position and this position is substituted by the less bulky prenyl group in molecule **6B**. In comparison with wild-type inhibition, mutants E474Q, F475A, and R482L have 5- to 10-fold less inhibition potency by both **2A** and **6B** indicating that these residues make specific contacts with the indole and quinone groups. A Cdc25B construct lacking the last 18 C-terminal residues (549–566) is five-fold less inhibited by **6B**, but has no significant change when inhibited by **2A**, suggesting that specific contacts are not established with the C-terminus by the indole and quinone groups. Mutants E478Q, Y528F, M531A, N532A, and R544L do not show significant changes when inhibited by **6B**, suggesting that these residues do not make specific contacts with ligands similar to **2**. Kinetic data for indolyldihydroxyquinones indicate that these are reversible and competitive ligands with time-independent inhibition.[70]

The above discussion may help to narrow down the possible binding modes found for **2**. Complex 2Ap is not a good model because specific contacts are established between the indole ring and the protein C-terminal. Complex 2Am seems unlikely because it makes no contact with R482 (or the neighbor P481), and coordinates residues which deletion or mutation does not affect binding considerably (W550, N532, and M531). Complex 2Dm is also unlike because it shows specific contacts with N532 and the 2-methyl-benzyl group is exposed to solvent. Similar reasoning suggests that 2Cz, 2Dz, 2Gz, and 2Hz are also unlikely.

The rigid Cdc25B crystal structure was used as a receptor model in another docking study.[78] For **1**, the binding mode found resembles complex 1Cp with the ligand slightly dislocated and forming a bidentate hydrogen bond with R482. Two modes were found for **2** corresponding to the same binding site as 2Ap, but with either 2-methyl-benzyl or dihydroxyquinone buried in the P-loop. The main difference in comparison with the results presented here were contacts observed between R544 and both **1** and **2**.[78] Given the same receptor structure was used, this variance is due to the different docking algorithms used.

CONCLUSIONS

Results from molecular dynamics simulations and bioinformatic searches presented here and of computer-generated conformational ensembles[43] indicate that the C-terminus is either unfolded or in equilibrium between a partially folded and a disordered state in Cdc25B under normal conditions in aqueous solution. Because of the growing interest in targeting Cdc25 phosphatases for development of anti-cancer therapeutics,[1–7] these results should warn those using the available Cdc25 crystal structures in structure-based drug design that such receptor configurations may be flawed or incomplete.

Complexes were obtained between three Cdc25 structural models and two quinone-based small molecules, which are potent Cdc25 inhibitors. The structural models had variable levels of flexibility so that the effect of Cdc25 C-terminal unfolding on small-molecule complexation could be evaluated. The same two binding sites were observed for **1** complexed with Cdc25B in the three structural models analyzed in agreement with experimentally measured mixed kinetics.[69] Binding to the shallow pocket formed beside the P-loop may account for irreversible oxidation of the catalytic C473.[76] Backbone flexibility has little influence on this binding site, but the contacts formed by **1** depend on the flexibility of P-loop and shallow pocket side-chains. Binding to the P-loop my occur in several orientations, which are clearly influenced by C-terminus backbone flexibility, and by the appearance of cavities unexposed in the rigid model. Complexation to the P-loop has a more favorable free

energy and may account for both irreversible oxidation and competitive inhibition.

In contrast to **1**, there should be only one mechanism of inhibition by **2** binding competitively to the active site.[70] This study and previously available data do not allow an unambiguous identification of this binding mode. The most likely possibilities are either described by complexes 2Cm and 2Em or 2Az and 2Bz or 2Fz. However, it is unlike that binding of **2** is a multimodal process.[64,75] Direct interactions with the C-terminus should be negligible for all complexes, but 2Fz. Nevertheless, inclusion of flexibility is essential to uncover the hydrophobic regions for binding the bulky 2-methyl-benzyl group and to obtain complexes in agreement with experimental data. It is expected that site-directed mutagenesis of Cdc25B residues D392-H395, Y400, Y428, P444-L445, K509-G510, and K513-E514 should help to resolve the occupation of the shallow pocket beside the P-loop and similar hydrophobic cavities.

As noted, backbone fluctuations alter the available binding sites in Cdc25B. The MD ensemble only hints at the Cdc25B transient cavities because of insufficient backbone sampling. A major advantage of using an ensemble following an equilibrium distribution, such as the Zuckerman ensemble as a structural model for receptor docking is the appropriate sampling of large scale fluctuations,[43] which help to disclosure the transient binding sites. The distribution of side chain rotamers, specially for side chain exposed to the solvent and inside binding pockets is also fundamental for ligand-receptor interactions.[23,26] But, because the side chain distribution in the Zuckerman ensemble does not follow an equilibrium distribution and given the lack of experimental structural information for comparison, we have refrained from discussing detailed interactions. It may be anticipated that ligands bound to Cdc25 will shift the population of side chain rotamers, as observed in the Cdc25A active site between apo[12] and substrate-bound structures.[79]

Contacts observed between ligands and putative disordered regions (here the Cdc25B C-terminal) in rigid receptor models will not be maintained when backbone flexibility is introduced because such disordered regions fluctuate wildly and do not assume a unique or stable set of configurations. New or cryptic binding sites involving the disordered region are not observed for the same reason. The main effect of modeling backbone flexibility is the formation of transient cavities or compact hydrophobic units[66] on the surface of the stable, folded protein core that are unexposed or unavailable for ligand binding in rigid and densely packed structures. These cavities are particularly important for binding to bulkier ligands with hydrophobic moieties. Repositioning of flexible but still ordered regions (e.g., residues M531-N532 in Cdc25B) also creates additional sites for contacts with ligands in comparison to a rigid model. Thus, the increased flexibility results in a larger diversity of simulated binding modes, in particular to bulkier ligands. Further, an approximate and economic receptor model for Cdc25B and other partially disordered proteins could be built by removing or truncating the partially unfolded or disordered regions (residues 531–550 for Cdc25B) from the available crystallographic or solution structure. Corroboration of the present conclusions and simulation results by structural and binding studies for different Cdc25B constructs is now underway in collaboration with experimental groups.

ACKNOWLEDGMENTS

Comments on the manuscript by Shaker Chuck Farah (Universidade de Sao Paulo) are acknowledged.

REFERENCES

1. Boutros R, Lobjois V, Ducommun B. Cdc25 phosphatases in cancer cells: key players? Good targets? Nat Rev Cancer 2007;7:495–507.
2. Brezak MC, Kasprzyk P, Galcera MO, Lavergne O, Prévost GP. Cdc25 inhibitors as anticancer agents are moving forward. Anti-Cancer Agents in Med Chem 2008;8:857–862.
3. Lazo JS, Wipf P. Is Cdc25 a druggable target? Anti-Cancer Agents in Med Chem 2008;8:837–842.
4. Bakan A, Lazo JS, Wipf P, Brummond KM, Bahar I. Toward a molecular understanding of the interaction of dual specificity phosphatases with substrates: insights from structure-based modeling and high throughput screening. Curr Med Chem 2008;19:2536–2544.
5. Vintonyak VV, Antonchick AP, Rauh D, Waldmann H. The therapeutic potential of phosphatase inhibitors. Curr Opin Chem Biol 2009;13:272–283.
6. Lavecchia A, Giovanni CD, Novellino E. Cdc25A and B dual-specificity phosphatase inhibitors: potential agents for cancer therapy. Curr Med Chem 2009;16:1831–1849.
7. Park H, Bahn YJ, Ryu SE. Structure-based de novo design and biochemical evaluation of novel Cdc25 phosphatase inhibitors. Bioorg Med Chem Lett 2009;19:4330–4334.
8. Rudolph J. Cdc25 phosphatases: structure, specificity, and mechanism. Biochemistry 2007;46:3595–3604.
9. Arantes GM. The catalytic acid in the dephosphorylation of the Cdk2-pTpY/CycA protein complex by Cdc25B phosphatase. J Phys Chem B 2008;112:15244–15247.
10. McCain DF, Catrina IE, Hengge AC, Zhang ZY. The catalytic mechanism of Cdc25A phosphatase. J Biol Chem 2002;277:11190–11200.
11. Arantes GM. Free energy profiles for catalysis by dual-specificity phosphatases. Biochem J 2006;399:343–350.
12. Fauman EB, Cogswell JP, Lovejoy B, Rocque WJ, Holmes W, Montana VG, Piwnica-Worms H, Rink MJ, Saper MA. Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. Cell 1998;93:617–625.
13. Reynolds RA, Yem AW, Wolfe CL, Deibel MR, Chidester CG, Watenpaugh KD. Crystal structure of the catalytic subunit of Cdc25b required for G(2)/M phase transition of the cell cycle. J Mol Biol 1999;293:559–568.
14. Sohn J, Parks JM, Buhrman G, Brown P, Kristjansdottir K, Safi A, Edelsbrunner H, Yang W, Rudolph J. Experimental validation of the docking orientation of Cdc25 with its Cdk2-CycA protein substrate. Biochemistry 2005;44:16563–16573.
15. Cavanagh J, Fairbrother WJ, Palmer AG, III, Skelton NJ. Protein NMR Spectroscopy: principles and Practice. 2nd edn. New York: Academic Press; 2007.
16. Wilborn M, Free S, Ban A, Rudolph J. The c-terminal tail of the dual-specificity Cdc25B phosphatase mediates modular substrate recognition. Biochemistry 2001;40:14200–14206.

17. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. The roles of intrinsic disorder in protein interaction networks. FEBS Lett 2005;272:5129–5148.
18. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of $p21^{Waf1/CiP1/Sdi1}$ in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci USA 1996;93:11504–11509.
19. Wright PE, Dyson HJ. Linking folding and binding. Curr Op Struct Biol 2009;19:31–38.
20. Jorgensen WL. The many roles of computation in drug discovery. Science 2004;303:1813–1818.
21. Sotriffer CA, Flader W, Winger RH, Rode BM, Leidl KR, Vargas JM. Automated docking of ligands to antibodies: methods and applications. Methods 2000;20:280–291.
22. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol 2008; 18:178–184.
23. Cozzini P, Kellogg GE, Spyrakis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M, Sotriffer CA. Target flexibility: an emerging consideration in drug discovery and design. J Med Chem 2008;51:6237–6255.
24. Kallblad1 P, Dean PM. Efficient conformational sampling of local side-chain flexibility. J Mol Biol 2003;326:1651–1665.
25. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in autodock. Proteins: Struct Funct Genet 2002;46:34–40.
26. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. J Med Chem 2004;47:45–55.
27. Knegtel RMA, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol 1997;266:424–440.
28. Prasad JC, Goldstone JV, Camacho CJ, Vajda S, Stegeman JJ. Ensemble modeling of substrate binding to cytochromes p450: analysis of catalytic differences between cyp1a orthologs. Biochemistry 2007;46:2640–2654.
29. Burra PV, Zhang Y, Godzik A, Stec B. Global distribution of conformational states derived from redundant models in the pdb points to non-uniqueness of the protein structure. Proc Natl Acad Sci USA 2009;106:10505–10510.
30. Damm KL, Carlson HA. Exploring experimental sources of multiple protein conformations in structure-based drug design. J Am Chem Soc 2007;129:8225–8235.
31. Davis AM, Teague SJ, Kleywegt GJ. Application and limitations of x-ray crystallographic data in structure-based ligand and drug design. Angew Chem Int Ed 2003;42:2718–2736.
32. Wong CF, Kua J, Zhang Y, Straatsma T, McCammon JA. Molecular docking of balanol to dynamics snapshots of protein kinase A. Proteins: Struct Funct Genet 2005;61:850–858.
33. Frembgen-Kesner T, Elcock AH. Computational sampling of a cryptic drug binding site in a protein receptor: explicit solvent molecular dynamics and inhibitor docking to p38 map kinase. J Mol Biol 2006;359:202–214.
34. Ensing B, Vivo MD, Liu Z, Moore P, Klein ML. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. Acc Chem Res 2006;39:73–81.
35. van Aalten D, de Groot B, Berendsen H, Findlay J, Amadei A. A comparison of techniques for calculating protein essential dynamics. J Comput Chem 1997;18:169–181.
36. Zuckerman DM, Lyman E. A second look at canonical sampling of biomolecules using replica exchange simulation. J Chem Theory Comput 2006;3:1200–1202.
37. Nagasima T, Sugita Y, Mitsutake A, Okamoto Y. Generalized-ensemble simulations of spin systems and protein systems. Comp Phys Comm 2002;146:69–76.
38. de Groot B, van Aalten D, Scheek R, Amadei A, Vriend G, Berendsen H. Prediction of protein conformational freedom from distance constraints. Proteins: Struct Funct Genet 1997;29: 240–251.
39. Lange OF, Lakomek NA, Fars C, Schrder GF, Walter KFA, Becker S, Meiler J, Grubmller H, Griesinger C, de Groot BL. Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution. Science 2008;320:1471–1475.
40. Shehu A, Kavraki LE, Clementi C. Multiscale characterization of protein conformationl ensembles. Proteins 2009;76:837–851.
41. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc Natl Acad Sci USA 2005;102:17002–17007.
42. Das R, Baker D. Macromolecular modeling with rosetta. Annu Rev Biochem 2008;77:363–382.
43. Mamonov AB, Bhatt D, Cashman DJ, Ding Y, Zuckerman DM. General library-based monte carlo technique enables equilibrium sampling of semi-atomistic protein models. J Phys Chem B 2009; 113:10891–10904.
44. Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from nmr residual dipolar couplings. J Am Chem Soc 2009;131:17908–17918.
45. McQuarrie DA. Statistical Mechanics. 1st edn. New York: Harper and Row; 1976.
46. Hess B, Kutzner C, van der Spoel D, Lindahl E. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:435–447.
47. Jorgensen WL, Maxwell DS, 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.
48. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
49. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg E, Man O, Beckmann JS, Silman I, Sussman JL. A simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 2005;21:3435–3438.
50. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol 2004;337:635–645.
51. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. Optimizing intrinsic disorder predictors with protein evolutionary information. J Bioinf Comput Biol 2005;3:35–60.
52. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. Biophys J 2007;92:1439–1456.
53. Liu JS. Monte Carlo strategies in scientific computing. 1st ed. New York: Springer; 2001.
54. Canutescu AA, Shelenkov AA, Dunbrack RL. A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci 2001;12:2001–2014.
55. Huey R, Morris GM, Olson AJ, Goodsell DS. A semiempirical free energy force field with charge-based desolvation. J Comput Chem 2007;28:1145–1152.
56. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. Angew Chem Int Ed 1999;38:236–240.
57. Haykin SS. Neural networks — A comprehensive foundation. 2nd ed. Upper Saddle River, NJ: Prentice-Hall; 1999.
58. Åqvist J, Luzhkov VB, Brandsdal BO. Ligand binding affinities from MD simulations. Acc Chem Res 2002;35:358–365.
59. Pearlman DA, Rao G. Free energy calculations: methods and applications. In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer IHF, Schreiner PR, editors Encyclopedia of Computational Chemistry, Vol. 2, 1st ed. Chichester: Wiley; 1998. pp. 1036–1061.
60. Hansson T, Marelius J, qvist J. Ligand binding affinity prediction by linear interaction energy methods. J Comput Aid Mol Des 1998;12: 27–35.

61. Benedix A, Becker CM, de Groot BL, Caflisch A, Bckmann RA. Predicting free energy changes using structural ensembles. Nat Methods 2009;6:3–4.
62. Krzeminski M, Fuentes G, Boelens R, Bonvin AM. Minoes: a new approach to select a representative ensemble of structures in nmr studies of (partially) unfolded states. Application to δ 25-PYP. Proteins 2009;74:895–904.
63. Lange OF, Grubmuller H. Generalized correlation for biomolecular dynamics. Proteins: Struct Funct Genet 2006;62:1053–1061.
64. Eftink MR, Anusiem AC, Biltonen RL. Enthalpy-entropy compensation and heat capacity changes for protein-ligand interactions: general thermodynamic models and data for the binding of nucleotides to ribonuclease A. Biochemistry 1983;22:3884–3896.
65. Bakan A, Bahar I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. Proc Natl Acad Sci USA 2009;105:14349–14354.
66. Teague SJ. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2003;2:527–541.
67. Heringa J, Argos P. Strain in protein structures as viewed through nonrotameric side chains: II. Effects upon ligand binding. Proteins: Struct Funct Genet 1999;37:44–55.
68. Gunasekaran K, Nussinov R. How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. J Mol Biol 2007;365:257–273.
69. Lazo JS, Aslan DC, Southwick EC, Cooley KA, Ducruet AP, Joo B, Vogt A, Wipf P. Discovery and biological evaluation of a new family of potent inhibitors of the dual specificity protein phosphatase Cdc25. J Med Chem 2001;44:4042–4049.
70. Sohn J, Kiburz B, Li Z, Deng L, Safi A, Pirrung MC, Rudolph J. Inhibition of Cdc25 phosphatases by indolyldihydroxyquinones. J Med Chem 2003;46:2580–2588.
71. Kallblad P, Mancera RL, Todorov NP. Assessment of multiple binding modes in ligand-protein docking. J Med Chem 2004;47:3334–3337.
72. Shortle D, Simons KT, Baker D. Clustering of low-energy conformations near the native structures of small proteins. Proc Natl Acad Sci USA 1998;95:11158–11162.
73. Krol M, Tournier AL, Bates PA. Flexible relaxation of rigid-body docking solutions. Proteins: Struct Funct Genet 2007;68:159169.
74. Rizzo RC, Tirado-Rives J, Jorgensen WL. Estimation of binding affinities for hept and nevirapine analogues with HIV-1 reverse transcriptase via Monte Carlo simulations. J Med Chem 2001;44:145–154.
75. Khandelwal A, Lukacova V, Kroll DM, Raha S, Comez D, Balaz S. Processing multimode binding situations in simulation-based prediction of ligand-macromolecule affinities. J Phys Chem A 2005; 109:6387–6391.
76. Brisson M, Nguyen T, Wipf P, Joo B, Day BW, Skoko JS, Schreiber EM, Foster C, Bansal P, Lazo JS. Redox regulation of Cdc25B by cell-active quinolinediones. Mol Pharmacol 2005;68:1810–1820.
77. Moser CC, Keske JM, Warncke K, Farid RS, Dutton PL. Nature of biological electron transfer. Nature 1992;355:796–802.
78. Lavecchia A, Cosconati S, Limongelli V, Novellino E. Modeling of Cdc25B dual specifity protein phosphatase inhibitors: docking of ligands and enzymatic inhibition mechanism. Chem Med Chem 2006;1:540–550.
79. Kolmodin K, Åqvist J. Prediction of a ligand-induced conformational change in the catalytic core of Cdc25A. FEBS Lett 2000; 465:8–11.