Article

Characterization of Folding Cores in the Cyclophilin A-Cyclosporin A Complex

Jack W. Heal,[1,2,*] Stephen A. Wells,[7] Claudia A. Blindauer,[4] Robert B. Freedman,[6] and Rudolf A. Römer[3,5]

[1]MOAC Doctoral Training Centre, [2]Institute for Advanced Study, [3]Centre for Scientific Computing, [4]Department of Chemistry, [5]Department of Physics, and [6]School of Life Sciences, University of Warwick, Coventry, United Kingdom; and [7]Department of Chemistry, University of Bath, Bath, United Kingdom

ABSTRACT   Determining the folding core of a protein yields information about its folding process and dynamics. The experimental procedures for identifying the amino acids that make up the folding core include hydrogen-deuterium exchange and Φ-value analysis and can be expensive and time consuming. Because of this, there is a desire to improve upon existing methods for determining protein folding cores theoretically. We have obtained HDX data for the complex of cyclophilin A with the immunosuppressant cyclosporin A. We compare these data, as well as literature values for uncomplexed cyclophilin A, to theoretical predictions using a combination of rigidity analysis and coarse-grained simulations of protein motion. We find that in this case, the most specific prediction of folding cores comes from a combined approach that models the rigidity of the protein using the FIRST software suite and the dynamics of the protein using the FRODA tool.

INTRODUCTION

The protein-folding problem has been a prevalent issue for the past 50 years, as the emerging protein structure crucially determines flexibility, mobility, and, ultimately, function (1). The two principal competing theories on how protein folding initiates are diffusion-collision (2) and nucleation-condensation (3). Indeed, it may well be that both are valid depending on which protein is being investigated (4). It can be intuited that residues that "collapse early during folding" (5) might be particularly important to the overall folding process and are usually referred to as defining a folding core. However, this set of residues is difficult to ascertain precisely. One way to define a folding core experimentally is through Φ-value analysis (6). This approach focuses on the folding process by using point mutations to determine the impact of particular residues on the energy of the transition state in a one-step folding process. An alternative is to study the dynamics of the folded structure through hydrogen-deuterium exchange (HDX) NMR experiments. For the examples of barnase and chymotrypsin inhibitor 2, it has been shown that the two definitions are consistent in that slowly exchanging residues in HDX have high Φ-values (7).

Establishing folding cores through Φ-value analysis or HDX provides valuable insight into protein folding and dynamics but also involves extensive experimental work. For this reason, the prediction of HDX folding cores through rapid computational methods is of ongoing interest (4,8–12). One method for predicting HDX folding cores uses rigidity analysis and is implemented in the FIRST software package (4,8). The method makes inferences about protein flexibility based solely upon the static crystal input structure. Surface exposure of the exchanging protons and protein motion, known to be important for HDX measurements and hence for folding cores (5,7), are ignored in this method. Here, we seek to improve upon the predictive power of rigidity analysis for folding-core prediction by incorporating missing information on surface exposure and dynamics of the protein.

We have selected cyclophilin A (CypA), a multifunctional 18 kDa protein with 165 residues, as the basis protein for our study, since it is large enough to exhibit complex folding behavior and at the same time readily investigated by HDX. It is known to bind strongly to the immunosuppressant drug cyclosporin A (CsA) (13–15). The structure of the CypA-CsA complex is shown in Fig. 1 with the binding-site residues highlighted (16).

In our experiment, we study the HDX behavior of unbound CypA and also its complex with CsA. Using these HDX data, we establish the resulting folding cores. Having found the HDX folding core in our experiments, we next apply FIRST to unbound CypA and to the CypA-CsA complex, establishing the theoretical FIRST folding core (FIR) in each case. We compare the resulting predicted folding cores with the HDX results and show that the established theoretical method of rigidity analysis implemented using the FIRST software provides a reasonable prediction of the experimental folding core (4,8,18). Nevertheless, although the FIRST method is impressive, it is not perfect in that it does not capture the changes in the HDX folding core
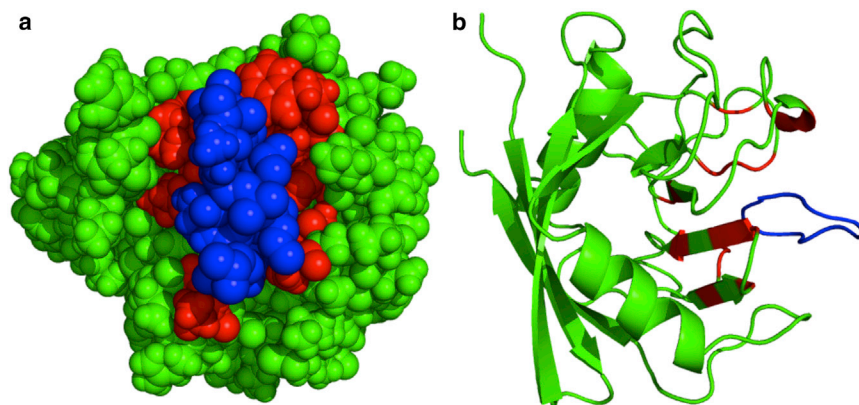
FIGURE 1 Sphere (*a*) and cartoon representation (*b*) of the CypA-CsA complex and the CypA binding site (using PDB structure 1CWA and the PyMOL visualizer (17)). CsA is indicated in blue and CypA in green. The 15 residues of CypA that have a heavy (nonhydrogen) atom within 4 Å of the atoms in CsA are red. To see this figure in color, go online.

observed upon ligand binding. Indeed, the predicted folding core is larger in the absence of the ligand, a result in contrast with what is observed experimentally.

We then change the hydrogen bond (HB) and hydrophobic tether (HP) networks so that these interactions are absent for surface atoms. With this modification, we enhance the flexibility of the protein surface so that the remaining rigid residues can better correlate with the experimental folding core. Building upon this, we model the prospective dynamics of the protein using coarse-grained simulations. By applying these techniques, we are able to probe the propensity for large-amplitude motions that can only be accessed over long timescales such as those of our HDX experiments. We show that our theoretical approach, which combines rigidity-analysis information from coarse-grained simulations and surface exposure, markedly improves the correlation between theory and experiment.

MATERIALS AND METHODS

The CypA-CsA complex

The multifunctional CypA belongs to the large class of ligand-binding proteins. CypA acts as a peptidyl-prolyl *cis-trans* isomerase in addition to performing other roles when binding to different molecules, such as the HIV-1 capsid protein (19–21). The CypA-CsA complex is strongly bound (13–15) with dissociation constant $K_D = 46$ nM (22). Most commonly used to suppress organ rejection after a transplant, CsA has also been administered to treat ulcerative colitis, cardiac disease, and a number of autoimmune diseases (23–25). It is the CypA-CsA complex that binds to and inhibits the T-cell activator calcineurin (CN) and thus has an immunosuppressant effect (26,27). Details of the expression and purification process and the characterization of the CypA-CsA complex are given in the Supporting Material.

HDX experiments have been conducted previously on unbound CypA (28) but not yet on the CypA-CsA complex. NMR has been used to solve the structure of CypA (15), as well as that of the CypA-CsA complex (13,14). Our HDX results broadly agree with the data on unbound CypA and in addition elucidate the effect of ligand binding in the CypA-CsA complex.

Rigidity analysis with FIRST

Protein rigidity analysis is a computational method that rapidly identifies rigid and flexible regions in a protein crystal structure (29,30). The structure

is considered as a molecular framework in which bond lengths and angles are fixed while dihedral angles are permitted to vary. Atomic degrees of freedom are then matched against bonding constraints (30–36). Covalent bonds, polar interactions (including hydrogen bonds and salt bridges), and hydrophobic tethers can all be included as bonding constraints. The output of the algorithm is a division of the structure into rigid clusters and flexible regions, known as a rigid-cluster decomposition (RCD). The RCD clearly depends on the constraints present in the bond network, the strength and location of which are determined solely from the geometry of the input structure. A systematic removal of hydrogen bonds from weakest to strongest leads to a loss in rigidity that we can relate to the unfolding of the protein (29). This is referred to as a rigidity dilution (RD). The pattern of rigidity loss can be used to gain insight into structural and functional properties of the protein (29,37,38). Indeed, RDs have been used previously to predict the HDX folding core for a number of proteins (4,8), not including CypA.

The Protein Data Bank (PDB) (39) x-ray crystal structure 1CWA of the CypA-CsA complex (40) was used for all simulations of protein rigidity and motion. After removing crystal water molecules, Reduce software (41) was used to add the hydrogen atoms and to flip side chains of Asn, Gln, and His residues where necessary. For simulations of the unbound protein, CsA was deleted from the structure manually using PyMOL (17), which was used for all molecular visualization. We note that the unbound structure for CypA is highly similar to the bound structure. Indeed, the backbone of 1CWA aligns with a different structure from that of the unbound protein (1W8V) with an RMSD of 0.26 Å.

The strength of each hydrogen bond, measured in kcal/mol, was calculated as a function of the geometry of the donor, hydrogen, and acceptor atoms using the distance- and angle-dependent Mayo potential (37,42). Only hydrogen bonds with a bond energy more negative than the energy cutoff parameter, E_{cut}, are included in the network. During an RD, E_{cut} is lowered, causing some hydrogen bonds to be excluded from the network. Rigidity dilution involves systematically lowering E_{cut} and reevaluating the RCD. Rigidity analysis was conducted using FIRST version 6.1 (43).

A modified bond network

We distinguish between buried and exposed residues in the protein by drawing points on the surface of a sphere around each atom, with radius equal to the van der Waals' radius of the atom plus that of a water molecule (1.4 Å). If the points of this sphere do not contact those of neighboring atoms, then the atom is labeled as being exposed. An atom is assigned a burial distance $R = 0$ Å if it is exposed, and $R = k$ Å otherwise, where k is the smallest distance to an exposed atom (see the Supporting Material for details).

We modify the standard bond network defined by FIRST based on the surface exposure of atoms with HBs and HPs. By demanding that both interacting atoms in an HB or HP constraint are buried within the protein, i.e., not exposed on the surface, we enhance the flexibility of the protein surface.

The folding core generated in this way consists of residues that are both rigid and buried within the protein.

Coarse-grained mobility simulation

The coarse-grained elastic network model implemented using the ElNemo software (44) allows for a rapid and accurate calculation of the normal modes of motion (45,46). The software FRODA allows rapid simulation of protein motion along these normal-mode vectors to generate new conformations satisfying the constraints of the FIRST bond network (37,47).

The 10 lowest-frequency nontrivial normal-mode vectors, m_7–m_{16}, capture well the large-scale motion of a protein (48). We use FRODA to generate 2000 conformers of the protein along both trajectories (parallel and antiparallel) defined by each of the vectors m_7–m_{16}. To generate a new conformer, we force each atom to move a distance of 0.01 Å in the direction specified by the normal-mode vector, as well as a distance of 0.01 Å in a random direction. We then check for steric clashes and impose the bond network determined by FIRST with $E_{cut} = -2.0$ kcal/mol, demanding that the bonds in this network are satisfied in terms of their distance and angle. We repeated these simulations at five different E_{cut} values (−0.5, −1.0, −1.5, −2.0, and −3.0 kcal/mol) and found that the nature of the results is not particularly sensitive to changes in this value (data not shown; see Heal (49)).

Quantitative measures for comparing folding cores

Let N_{Ex} and N_{Th} be the numbers of residues contained in an experimentally and theoretically determined folding core, respectively. Clearly, $N_{Ex} = N_{Th}$ is a necessary condition for agreement of experimentally and theoretically estimated folding cores. However, it is not just the number of residues, but of course the agreement of the specific set of residues in both experimental and theoretical folding cores, that is is most important. To capture this, let us define $\mathcal{T}$ as the number of residues correctly identified by a theoretical prediction of the experimental folding core, i.e., $\mathcal{T}$ is the number of true positives. For a perfect agreement, we have $\mathcal{T} = N_{Th}$ while the expected $\mathcal{T}$ attained randomly is $N_{Th} N_{Ex} / N$. Here, N denotes the total number of amino acids in the protein (8). We can define the specificity, α, and sensitivity, γ, of a theoretical folding core prediction as

$$\alpha = \frac{\mathcal{T}}{N_{Th}} \qquad (1)$$

and

$$\gamma = \frac{\mathcal{T}}{N_{Ex}}. \qquad (2)$$

The specificity, α, measures the fraction of residues identified by the theoretical method that are also part of the experimental folding core. The sensitivity, γ, shows the proportion of residues in the experimental core that have been correctly predicted by the theoretical method. A perfect correspondence between theory and experiment, $\mathcal{T} = N_{Th} = N_{Ex}$, would yield $\alpha = \gamma = 1$, whereas for a completely wrong identification, $\mathcal{T} = 0$, we have $\alpha = \gamma = 0$.

Another, previously defined, quantitative measure (8) is the so-called folding-core identification enhancement factor, ϵ. This is the ratio of $\mathcal{T}$ to the number of residues expected to be identified by a random selection,

$$\epsilon = \frac{\mathcal{T} N}{N_{Th} N_{Ex}}. \qquad (3)$$

A theoretical method with random probability of success has $\epsilon = 1$, and when $\epsilon > 1$, the match is better than random. We shall also compare our α and γ measures to ϵ.

Experimental procedure

The following is a brief outline of the experimental procedure. Full details of protein expression, purification, biophysical characterization, NMR spectrum assignments, and HDX experiments are presented in the Supporting Material. The protocol for protein expression and purification was adapted from Liu et al. (22). Circular dichroism was used to determine the folded state of the protein and fluorescence spectroscopy to demonstrate CypA-CsA complex formation (50,51). NMR assignments were determined from a series of 3D (^{1}H,^{15}N,^{1}H) total correlation spectroscopy-heteronuclear single quantum coherence (TOCSY-HSQC) and nuclear Overhauser effect spectroscopy-heteronuclear single quantum coherence (NOESY-HSQC). For the HDX experiments, lyophilized protein was resuspended in phosphate buffer at pH 6.5 prepared with D$_2$O. Exchange was monitored through a series of two-dimensional [^{1}H,^{15}N] HSQC spectra acquired on a Bruker AV II 700 spectrometer.

RESULTS AND DISCUSSION

The FIRST folding cores

We simulated RDs by systematically lowering E_{cut}, i.e., removing the hydrogen bonds in order of strength from weakest to strongest (37). We used FIRST to generate an RCD each time a hydrogen bond was removed. In a one-dimensional representation of an RCD, each residue in the primary structure is labeled as being rigid or flexible depending on the rigidity of its C_α atom. We show rigid residues as blocks colored according to their rigid-cluster membership.

We visualize the pattern of rigidity loss during RDs by plotting the one-dimensional representation of the RCD each time this changes. Such plots for the CypA-CsA complex and the unbound CypA are given in Fig. 2. When $|E_{cut}|$ is small, the protein is largely rigid and many of the residues are represented as blocks. As E_{cut} becomes more negative, i.e., as stronger bonds are excluded from the bond network, more residues become flexible.

In both of the RD plots, there is a clear and abrupt transition from the largely rigid state to the largely flexible state, consistent with the first-order rigidity loss expected for a predominantly β-sheet protein (37). The lowest line in the RD plot, where at least three residues of two or more secondary structures (as determined using the DSSP algorithm (52)) are part of the same rigid cluster, determines the FIRST folding core (4,8,18,53). We refer to the E_{cut} corresponding to this line as the folding-core energy, E_{fc}. For unbound CypA, $E_{fc} = -1.263$ kcal/mol, and for the CypA-CsA complex, $E_{fc} = -1.452$ kcal/mol. That the CypA-CsA complex has a lower E_{fc} suggests that ligand binding confers stability on the complex, as more bonds need to be broken to render the protein mostly flexible (53). The residues that are mutually rigid in the RCD evaluated at the E_{fc} (Fig. 2, red), form the FIRST folding core.

The HDX folding cores

We define the set of residues for which a corresponding backbone amide signal remains in the HSQC spectrum after
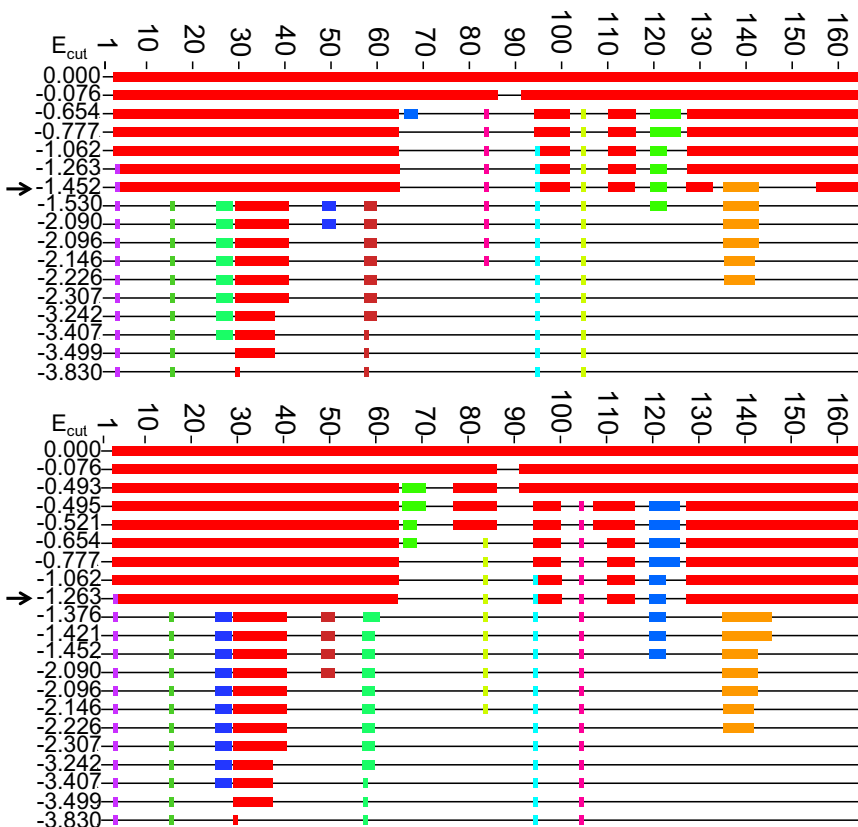
FIGURE 2 RD plots of the CypA-CsA complex (*upper*) and unbound CypA (*lower*). The RCD is shown at different values of E_{cut} (kcal/mol). Rigid residues are shown as thick colored blocks and flexible regions as thin horizontal black lines. Residues that are mutually rigid are shown in the same color. The line representing the FIRST folding core in each case is indicated by an arrow. To see this figure in color, go online.

110 min as the experimentally determined HDX folding core of the complex. These residues are listed in Table 1 and are also indicated in Fig. 3.

The published HDX experiments on unbound CypA resulted in a classification of CypA residues in terms of their exchange rates, k_{ex} (28). Twelve residues, including the proline residues, were not categorized, since they were not identified in the HSQC spectrum. We also carried out HDX experiments on the unbound protein, and our data, shown in the Supporting Material, were in agreement with the previously published result (28).

Here, we have drawn on the exchange rates determined in Shi et al. (28) and defined the residues with $k_{ex} < 10^{-2} \, \text{min}^{-1}$ as the folding core of the unbound protein, a definition congruent with that applied to our data set for the CypA-CsA complex (see above). According to this approach, the HDX folding core for unbound CypA has 73 residues.

Comparison of FIRST and HDX folding cores

In Fig. 3, we compare the HDX and FIRST folding cores along the primary structure of the CypA-CsA complex and unbound CypA. Residues that form part of the folding cores are represented as colored blocks. In both cases, the FIRST folding cores largely overlap with the HDX folding cores. The small changes that do occur upon ligand binding in the HDX folding cores are not very well captured in the

two FIRST folding cores. Rather, these differ only between residues 133 and 155, whereas the HDX folding cores remain largely unaffected in this region. Notably, the FIRST folding core for the CypA-CsA complex is smaller than that for the unbound protein. This contrasts with both our expectation and experimental finding as given above, where we show that the HDX folding core increases in size upon ligand binding. It highlights a problem with a theoretical method that only uses FIRST. Ligand binding to the surface of CypA causes the binding-site residues to become buried where before they were exposed, which may affect their HDX exchange rates. This effect is not modeled in FIRST, where we merely consider the hydrogen-bond network of the initial, static crystal structure.

TABLE 1 Residues of the HDX folding core for the CypA-CsA complex

CypA-CsA HDX folding-core residues
V6, F7, F8, D9, I10, A11, V12, E15, L17, V20, **S21**, F22, E23, L24, F25, A26, V29, K31, T32, A33, E34, N35, F36, R37, A38, L39, S40, T41, Y48, S51, F53, **H54**, R55, I57, F60, M61, Q63, G64, **K76, I78**, E86, **N87, F88, I89**, L90, G96, I97, L98, S99, M100, **N108**, F112, F113, I114, C115, T116, A117, **T119, L122**, D123, K125, V127, V128, F129, G130, K131, V132, K133, I138, V139, E140, A141, M142, E143, F145, T157, A159, D160, G162

Residues in bold print are slowly exchanging only in the presence of the ligand.
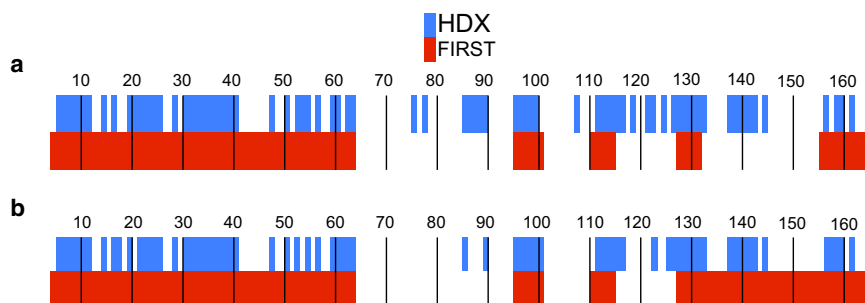
FIGURE 3 The HDX folding cores (*blue*) are given along with the FIRST folding cores (*red*) for the CypA-CsA complex (*a*) and unbound CypA (*b*). The residue numbers along the protein backbone are indicated above each section, with thin vertical lines added every 10 residues for clarity. Only the residues that are part of the folding cores are colored. To see this figure in color, go online.

We find that the HDX folding core does not change dramatically upon ligand binding. The 10 residues in bold in Table 1 are slowly exchanging in the presence of the CsA ligand but not part of the HDX folding core for unbound CypA. Eight of these belong to flexible regions proximal to the binding site. These changes suggest shielding or constraining of the unstructured regions near the binding site in the presence of the ligand. The small increase in folding-core size, from 73 to 80 residues, is consistent with the expectation that the presence of a ligand is likely to shield certain residues from the surface of the protein. There are also three assigned residues—G18, A101, and I158—that are, surprisingly, part of the unbound folding core according to Shi et al. (28) but not part of the bound folding core.

Folding cores FIR$_B$, SeS, FRO, and F+F

We now introduce four additional theoretical folding cores, which we consider alongside the FIRST folding core in our subsequent analysis. We refer to these folding cores as FIR$_B$, SeS, FRO, and F+F. FIR$_B$ is similar to the FIRST folding core, but with the additional condition that a residue must be buried as well as rigid to be part of FIR$_B$. This condition was imposed through the modified bond network in FIRST, which demands that both partners in an HB or HP network are buried.

For SeS, we take the secondary structural units (α-helix and β-strand) as determined using the DSSP algorithm.

This is our simplest theoretical predictor of a folding core based only on secondary structure. The results of our coarse-grained simulations of protein motion are incorporated in FRO. Here, the final (2000th) conformer from each of our 20 simulations was subjected to a burial-distance analysis, and those residues that were buried in >10 of these conformers were included in FRO. The burial distance, R, of each residue was determined from its amide nitrogen (see the Supporting Material), and if $R > 1.5$ Å, the residue was classed as buried. We emphasize that the results of this mobility-based analysis depend indirectly on the results of the rigidity analysis; FIRST identifies a constraint network and FRODA then explores the motion that is possible within those constraints. Therefore, it is possible for a residue to be 1) rigid but exposed, thus lying in FIR or FIR$_B$ but not in FRO; or 2) nonrigid but well protected by burial within the protein, thus lying in FRO but not in FIR or FIR$_B$. This suggests that to predict the results of the HDX folding core, which depend upon surface exposure as well as dynamics (inherited from flexibility), a combination of the approaches employed by FIRST and FRODA may be necessary. F+F is simply the intersection of residues in FIR and FRO.

In Fig. 4, each of the five theoretical folding cores is compared with the HDX folding core for the CypA-CsA complex. In each case, folding-core residues along the protein backbone are shown as bold blocks of color.

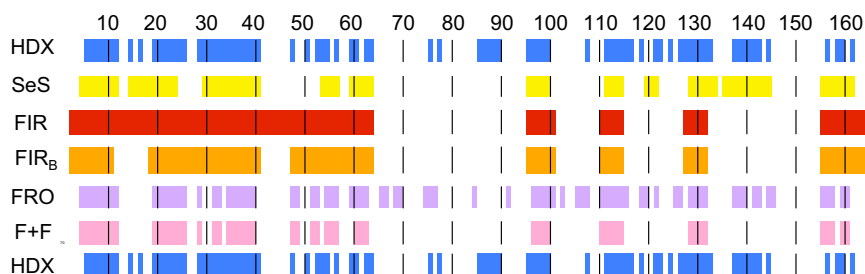We find that SeS consists almost entirely of slowly exchanging amino acids, although there are also many

FIGURE 4 Comparison of the experimentally determined HDX folding core for CypA-CsA (*blue*; shown both above and below the other folding cores), with the five folding cores computed using the theoretical approaches outlined in the text and labeled accordingly (from top to bottom, SeS (*yellow*), FIR (*red*), FIR$_B$ (*orange*), FRO (*purple*), and F+F (*pink*)). In each case, the folding-core residues in the primary structure of CypA are represented as colored blocks and the nonfolding core residues are white. Residue numbers along the protein backbone are indicated above the table body, with thin black vertical lines added every 10 residues for clarity. To see this figure in color, go online.

slowly exchanging residues that are not part of a secondary structure unit. As expected, FIR_B is smaller than FIR, since there are fewer constraints present throughout the rigidity dilution. The surface-exposed residues 12–18 and 41–47 are part of FIR but not part of FIR_B. The residues in FRO are also for the most part slowly exchanging. The combined F+F folding core matches the HDX core with high specificity. Fig. 5 presents a graphical comparison between the HDX folding core and F+F on the structure of CypA. We observe that in the absence of the ligand, residues 86–90 are mostly absent from both theoretical and experimental folding cores. The relatively small number of false positives (Fig. 5, *red*: 10 residues in (*a*) and 8 in (*b*)) relative to false negatives (Fig. 5, *blue*: 36 residues in (*a*) and 22 in (*b*)) demonstrate the high ratio of specificity to sensitivity for this method.

We note that some parts of the HDX folding core, for example, residues 86–90, are poorly predicted by all of the methods. This region corresponds to a flexible, surface-exposed, unstructured region in the crystal structure (Fig. 5 *a*, *arrow*). Our methods, based around predicting rigid, buried, and immobile regions, fall short of identifying this area as slowly exchanging.

Quantitative analysis of theoretical folding cores

Table 2 shows the number of residues in each theoretical folding core, along with the number of true positives, $\mathcal{T}$. The comparison pictured in Fig. 4 is made quantitative with the specificity, α, sensitivity, γ, and enhancement factor, ϵ, as defined in Eqs. 1–3, respectively.

The ϵ values of the five theoretical folding cores, the lowest of which is 1.33 (for the previously established FIRST folding core FIR), demonstrate improvement over random selection. F+F scores highest for both the CypA-CsA complex and the unbound CypA, suggesting that the combined approach has merit. This method is highly specific ($\alpha \geq 0.8$ in both cases), although its low γ scores show that it does not capture enough of the HDX folding core to suc-

cessfully capture the impact of ligand binding on HDX. SeS is highly specific ($\alpha \geq 0.77$) and sensitive ($\gamma \geq 0.74$), although this method of folding-core selection is clearly inappropriate for predicting changes upon ligand binding, since it does not change. The two folding cores derived solely using FIRST (FIR and FIR_B) do not reflect the experimentally observed increase in folding-core size upon ligand binding. Only when dynamics are also incorporated, i.e., in FRO and F+F, is this qualitative effect reflected.

We have also estimated the variation in α and γ in Table 2, assuming a $\pm 5\%$ error in N_{Ex}, N_{Th}, and $\mathcal{T}$. The six proline residues in CypA do not appear in the $^{1}H,^{15}N$ HSQC spectrum due to their lack of an amide proton. For this reason, HDX experiments cannot inform on whether prolines are part of a folding core. We therefore also calculated ϵ with $N = 159$. The ordering of the theoretical methods remains the same, and coupled with the small variation in α and γ, this shows that our quantitative comparison is robust.

CONCLUSIONS

The experimental HDX folding core for the CypA-CsA complex is highly similar to that for the unbound protein, albeit with a small number of additional residues. This small change is consistent with previous observations of only subtle conformational change in CypA upon ligand binding (15,54). The FIRST folding cores differ more substantially. Ligand binding confers rigidity upon the structure but alters the pattern of rigidity loss so that the FIRST folding core in fact decreases in size. In both cases, the FIRST folding core is a reasonable match for the HDX folding core, in agreement with previous folding-core predictions using FIRST (4). The FIRST folding core is defined by the RCD at E_{fc}, a value that decreases upon ligand binding. This means that when the folding cores are compared for the protein before and after ligand removal, we are comparing RCDs at different E_{cut} values. As a result, there may be more
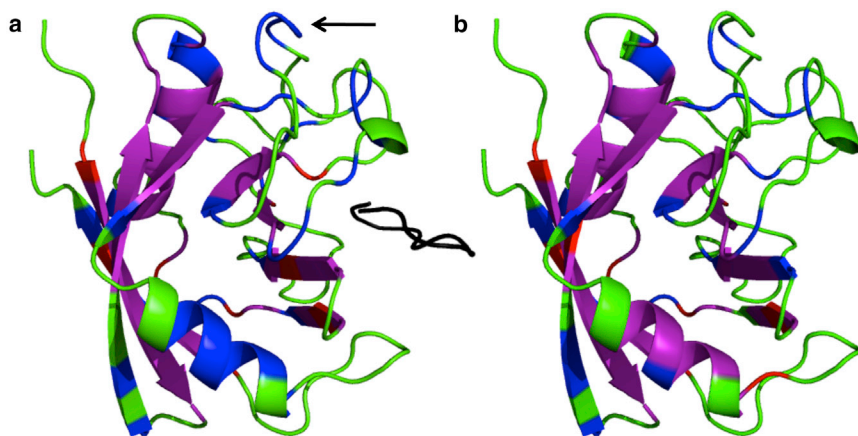

FIGURE 5 Cartoon representation of a comparison of HDX folding cores with F+F folding cores for the CypA-CsA complex (*a*) and unbound CypA (*b*). In both cases, residues that are part of both folding cores are purple, those found only in the HDX folding core are blue, those in the F+F folding core only are red, and those that are not part of either folding core are red. CsA is colored black in (*a*). The arrow in (*a*) indicates residues 86–90. To see this figure in color, go online.

TABLE 2 Quantitative measures of agreement between theoretical folding cores and the experimental HDX folding core

Folding core	N_{Ex}	N_{Th}	$\mathcal{T}$	α	γ	$\in_{165}$	$\in_{159}$
CypA-CsA							
SeS	80	75	59	0.79 ± 0.04	**0.74** ± 0.04	1.62	1.56
FIR	80	86	57	0.66 ± 0.04	0.71 ± 0.04	1.37	1.32
FIR$_B$	80	72	53	0.74 ± 0.04	0.66 ± 0.04	1.52	1.46
FRO	80	80	56	0.70 ± 0.04	0.70 ± 0.04	1.44	1.39
F+F	80	54	44	**0.81** ± 0.05	0.55 ± 0.03	**1.68**	**1.62**
Unbound CypA							
SeS	73	75	58	0.77 ± 0.04	0.79 ± 0.04	1.75	1.68
FIR	73	112	66	0.59 ± 0.03	**0.90** ± 0.05	1.33	1.28
FIR$_B$	73	72	52	0.72 ± 0.04	0.71 ± 0.04	1.63	1.57
FRO	73	76	54	0.71 ± 0.04	0.74 ± 0.04	1.61	1.54
F+F	73	59	51	**0.86** ± 0.05	0.70 ± 0.04	**1.95**	**1.88**

Values are for CypA-CsA complex (*upper*) and unbound CypA (*lower*), with $N = 165$. Bold numbers indicate the largest selectivity, α, sensitivity, γ, and enhancement, $\in$, for CypA-CsA and unbound CypA. The error estimates for α and γ reflect an assumed ± 5% variation in N_{Ex}, N_{Th}, and $\mathcal{T}$. The two estimates for the enhancement, $\in$, show the variation when using the full set of residues with $N = 165$ and with the six proline residues of CypA excluded such that $N = 159$.

constraints present in the E_{cut} value for the unbound protein than in that for the complex, resulting in the unlikely prediction of a larger folding core in the absence of a ligand. FIRST is a tool that can be implemented rapidly, and this study complements previous studies demonstrating its utility for folding-core prediction (4,8,18,53). Nevertheless, our work also shows that the FIRST-based folding-core predictions are not yet accurate or sensitive enough to capture the impact of ligand binding for CypA. For the purpose of predicting subtle effects of ligand binding, simply analyzing patterns of rigidity, as done through FIRST, appears insufficient. The trade-off between rapid computation on the one hand, as achieved with FIRST, and the necessary accuracy of folding core prediction on the other hand needs to be more finely balanced.

As a way forward, we have introduced and discussed rapid computational methods for adding information on surface exposure and protein dynamics into a rigidity-based folding core definition. These improve markedly upon the FIRST folding core for predicting the HDX folding cores of both CypA-CsA and unliganded CypA. In addition, we show that they are less sensitive to the erroneous increase in folding-core size upon ligand removal observed using FIRST with the standard bond network. On balance, F+F, which combines rigidity and motion, appears to be the best choice for a computational determination of the folding core from a protein's structural information alone. Nevertheless, no method achieves a perfect score of $\alpha = \gamma = 1$, either for the uncomplexed CypA or for the CypA-CsA complex. This insensitivity to ligand binding is disappointing but unsurprising due to the scale of the challenge, both for CypA-CsA and in general.

SUPPORTING MATERIAL

Supporting Materials and Methods, seven figures, and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00182-4.

REFERENCES

1. Dill, K. A., and J. L. MacCallum. 2012. The protein-folding problem, 50 years on. *Science.* 338:1042–1046.

2. Karplus, M., and D. L. Weaver. 1994. Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.* 3:650–668.

3. Itzhaki, L. S., D. E. Otzen, and A. R. Fersht. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* 254:260–288.

4. Hespenheide, B. M., A. J. Rader, …, L. A. Kuhn. 2002. Identifying protein folding cores from the evolution of flexible regions during unfolding. *J. Mol. Graph. Model.* 21:195–207.

5. Woodward, C. 1993. Is the slow exchange core the protein folding core? *Trends Biochem. Sci.* 18:359–360.

6. Fersht, A. R., A. Matouschek, and L. Serrano. 1992. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* 224:771–782.

7. Li, R., and C. Woodward. 1999. The hydrogen exchange core and protein folding. *Protein Sci.* 8:1571–1590.

8. Rader, A. J., and I. Bahar. 2004. Folding core predictions from network models of proteins. *Polymer (Guildf.).* 45:659–668.

9. Tartaglia, G. G., A. Cavalli, and M. Vendruscolo. 2007. Prediction of local structural stabilities of proteins from their amino acid sequences. *Structure.* 15:139–143.

10. Liu, T., D. Pantazatos, …, V. L. Woods, Jr. 2012. Quantitative assessment of protein structural models by comparison of H/D exchange MS data with exchange behavior accurately predicted by DXCOREX. *J. Am. Soc. Mass Spectrom.* 23:43–56.

11. Lobanov, M. Y., M. Y. Suvorina, …, O. V. Galzitskaya. 2013. A novel web server predicts amino acid residue protection against hydrogen-deuterium exchange. *Bioinformatics.* 29:1375–1381.

12. Thomas, S., L. Tapia, …, N. M. Amato. 2013. Rigidity analysis for protein motion and folding core identification. Paper from the AAAI 2013 Workshop.

13. Neri, P., R. Meadows, …, S. Fesik. 1991. ^1H, ^{13}C, and ^{15}N backbone assignments of cyclophilin when bound to cyclosporin A (CsA) and preliminary structural characterization of the CsA binding site. *FEBS Lett.* 294:81–88.

14. Spitzfaden, C., W. Braun, …, K. Wüthrich. 1994. Determination of the NMR solution structure of the cyclophilin A-cyclosporin A complex. *J. Biomol. NMR.* 4:463–482.

15. Ottiger, M., O. Zerbe, …, K. Wüthrich. 1997. The NMR solution conformation of unligated human cyclophilin A. *J. Mol. Biol.* 272:64–81.

16. Rader, A. J., and S. M. Brown. 2011. Correlating allostery with rigidity. *Mol. Biosyst.* 7:464–471.

17. Schrödinger, L. L. C. 2013. The PyMOL Molecular Graphics System, Version 1.3r1. http://www.pymol.org/pyMOL.

18. Rader, A. J., G. Anderson, …, J. Klein-Seetharaman. 2004. Identification of core amino acids stabilizing rhodopsin. *Proc. Natl. Acad. Sci. USA.* 101:7246–7251.

19. Luban, J., K. L. Bossolt, …, S. P. Goff. 1993. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell.* 73:1067–1078.

20. Bosco, D. A., and D. Kern. 2004. Catalysis and binding of cyclophilin A with different HIV-1 capsid constructs. *Biochemistry.* 43:6110–6119.

21. Wang, P., and J. Heitman. 2005. The cyclophilins. *Genome Biol.* 6:226.

22. Liu, J., M. W. Albers, …, C. T. Walsh. 1990. Cloning, expression, and purification of human cyclophilin in *Escherichia coli* and assessment of the catalytic role of cysteines by site-directed mutagenesis. *Proc. Natl. Acad. Sci. USA.* 87:2304–2308.

23. Nussenblatt, R. B., and A. G. Palestine. 1986. Cyclosporine: immunology, pharmacology and therapeutic uses. *Surv. Ophthalmol.* 31: 159–169.

24. Lichtiger, S., D. H. Present, A. Kornbluth, I. Gelernt, J. Bauer, G. Galler, F. Michelassi, and S. Hanauer. 1994. Cyclosporine in severe ucleraive colilitis refractory to steroid therapy. *New. Eng. J. Med.* 330:1841–1845.

25. Mott, J. L., D. Zhang, …, H. P. Zassenhaus. 2004. Cardiac disease due to random mitochondrial DNA mutations is prevented by cyclosporin A. *Biochem. Biophys. Res. Commun.* 319:1210–1215.

26. Zydowsky, L. D., F. A. Etzkorn, …, C. T. Walsh. 1992. Active site mutants of human cyclophilin A separate peptidyl-prolyl isomerase activity from cyclosporin A binding and calcineurin inhibition. *Protein Sci.* 1:1092–1099.

27. Liu, Y., J. Jiang, …, W. M. Kati. 2006. A fluorescence polarization-based assay for peptidyl prolyl *cis/trans* isomerase cyclophilin A. *Anal. Biochem.* 356:100–107.

28. Shi, Y.-H., D.-H. Lin, …, X. Shen. 2006. Study of structural stability of cyclophilin A by NMR and circular dichroism spectra. *Chin. J. Chem.* 24:973–979.

29. Rader, A. J., B. M. Hespenheide, …, M. F. Thorpe. 2002. Protein unfolding: rigidity lost. *Proc. Natl. Acad. Sci. USA.* 99:3540–3545.

30. Jacobs, D. J., A. J. Rader, …, M. F. Thorpe. 2001. Protein flexibility predictions using graph theory. *Proteins.* 44:150–165.

31. Jacobs, D. J., and M. F. Thorpe. 1995. Generic rigidity percolation: the pebble game. *Phys. Rev. Lett.* 75:4051–4054.

32. Jacobs, D. J., and B. Hendrickson. 1997. An algorithm for two-dimensional rigidity percolation: the pebble game. *J. Comput. Phys.* 137: 346–365.

33. Jacobs, D. J. 1998. Generic rigidity in three-dimensional bond-bending networks. *J. Phys. A Math. Gen.* 31:6653–6668.

34. Jacobs, D. J., L. A. Kuhn, and M. F. Thorpe. 1999. Flexible and rigid regions in proteins. *In* Rigidity Theory and Applications. M. F. Thorpe and P. M. Duxbury, editors. Kluwer Academic/Plenum, New York, pp. 357–384.

35. Thorpe, M. F., B. M. Hespenheide, …, L. A. Kuhn. 2000. Flexibility and critical hydrogen bonds in cytochrome *c*. *Pac. Symp. Biocomput.* 2000:191–202.

36. Hespenheide, B. M., D. J. Jacobs, and M. Thorpe. 2004. Structural rigidity and the capsid assembly of cowpea chlorotic mottle virus. *J. Phys. Condens. Matter.* 16:S5055–S5064.

37. Wells, S. A., J. E. Jimenez-Roldan, and R. A. Römer. 2009. Comparative analysis of rigidity across protein families. *Phys. Biol.* 6:046005–046011.

38. Heal, J. W., J. E. Jimenez-Roldan, …, R. A. Römer. 2012. Inhibition of HIV-1 protease: the rigidity perspective. *Bioinformatics.* 28:350–357.

39. The Protein Data Bank. 2013. http://www.rcsb.org/pdb/.

40. Mikol, V., J. Kallen, …, M. D. Walkinshaw. 1993. X-ray structure of a monomeric cyclophilin A-cyclosporin A crystal complex at 2.1 A resolution. *J. Mol. Biol.* 234:1119–1130.

41. Word, J. M., S. C. Lovell, …, D. C. Richardson. 1999. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* 285:1735–1747.

42. Dahiyat, B. I., D. B. Gordon, and S. L. Mayo. 1997. Automated design of the surface positions of protein helices. *Protein Sci.* 6:1333–1337.

43. Flexweb. 2013. Analysis of Flexibility in Biomolecules and Networks. http://flexweb.asu.edu/software/first/.

44. Suhre, K., and Y.-H. Sanejouand. 2004. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* 32:W610–W614, (Web Issue).

45. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from single-parameter atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.

46. Hinsen, K. 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins.* 33:417–429.

47. Wells, S., S. Menor, …, M. F. Thorpe. 2005. Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.* 2:S127–S136.

48. Jimenez-Roldan, J. E., R. B. Freedman, …, S. A. Wells. 2012. Rapid simulation of protein motion: merging flexibility, rigidity and normal mode analyses. *Phys. Biol.* 9:016008.

49. Heal, J. W. 2013. Effects of ligand binding on the rigidity and mobility of proteins: a computational and experimental approach. PhD thesis. University of Warwick, Coventry, United Kingdom.

50. Husi, H., and M. G. M. Zurini. 1994. Comparative binding studies of cyclophilins to cyclosporin A and derivatives by fluorescence measurements. *Anal. Biochem.* 222:251–255.

51. Gastmans, M., G. Volckaert, and Y. Engelborghs. 1999. Tryptophan microstate reshuffling upon the binding of cyclosporin A to human cyclophilin A. *Proteins.* 35:464–474.

52. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.

53. Tastan, O., E. Yu, …, J. Klein-Seetharaman. 2007. Comparison of stability predictions and simulated unfolding of rhodopsin structures. *Photochem. Photobiol.* 83:351–362.

54. Fanghänel, J., and G. Fischer. 2003. Thermodynamic characterization of the interaction of human cyclophilin 18 with cyclosporin A. *Biophys. Chem.* 100:351–366.