Docking of Photosystem I Subunit C Using a Constrained Geometric Simulation

ARTICLE *in* JOURNAL OF THE AMERICAN CHEMICAL SOCIETY · AUGUST 2006

Impact Factor: 12.11 · DOI: 10.1021/ja0587749 · Source: PubMed

5 AUTHORS, INCLUDING:

Craig C Jolley
RIKEN
35 PUBLICATIONS **392** CITATIONS
SEE PROFILE

Stephen Wells
University of Bath
66 PUBLICATIONS **1,147** CITATIONS
SEE PROFILE

Michael Thorpe
Arizona State University
279 PUBLICATIONS **9,795** CITATIONS
SEE PROFILE

Petra Fromme
Arizona State University
144 PUBLICATIONS **8,750** CITATIONS
SEE PROFILE

J|A|C|S
ARTICLES
Published on Web 06/16/2006

Docking of Photosystem I Subunit C Using a Constrained Geometric Simulation

Craig C. Jolley,[†,§,¶] Stephen A. Wells,[†,¶] Brandon M. Hespenheide,[†,¶] Michael F. Thorpe,[†,‡,¶] and Petra Fromme*[,‡,§]

Contribution from the Departments of Physics & Astronomy and Chemistry & Biochemistry, The Center for the Study of Early Events in Photosynthesis, Arizona State University, and The Center for Biological Physics, Biodesign Institute at Arizona State University, Tempe, Arizona 85287

Received January 9, 2006; E-mail: pfromme@asu.edu

Abstract: The elucidation of assembly pathways of multi-subunit protein complexes is a problem of great interest in structural biology and biomolecular modeling. In this study, we use a new computer algorithm for the simulation of large-scale motion in proteins to dock the subunit PsaC onto Photosystem I. We find that a complicated docking pathway involving multiple conformational changes can be quickly simulated by actively targeting only a few residues at a time to their target positions. Simulations for two possible docking scenarios are explored, and experimental approaches to distinguish between them are discussed.

Introduction

Photosystem I (PSI) is a large transmembrane protein complex that is vital for the process of oxygenic photosynthesis, the biological process by which solar energy is converted into the chemical energy used by all higher forms of life on Earth.[3,4] PSI catalyzes light-driven electron transfer across the photosynthetic membrane (Figure 1), from its lumenal (i.e., inside the thylakoid membrane) electron donor plastocyanin/cytochrome c_6 to its stromal (outside the thylakoid membrane) electron acceptor ferredoxin. In cyanobacteria, PSI forms a trimer in which each monomer contains 12 protein subunits and 127 noncovalently bound cofactors. More specifically, each monomer binds 96 chlorophyll *a* molecules, 21 β-carotenes, 2 phylloquinones, 4 lipids, 1 calcium cation, and 3 [4Fe-4S] clusters, which consist of 4 iron atoms and 4 sulfur atoms arranged at the corners of a cube. Of these 12 protein subunits, 9 of them (PsaA, PsaB, PsaF, PsaI, PsaJ, PsaK, PsaL, PsaM, and PsaX) are primarily α-helical and span the thylakoid membrane.

The other three subunits (PsaC, PsaD, and PsaE) do not contain transmembrane helices. Instead, they are soluble proteins which dock onto the stromal surface of the membrane-intrinsic portion of PSI, forming what is known as the stromal hump.[5] Solution structures for PsaC[6] and PsaE[7−9] have been obtained using NMR, while PsaD is believed to be largely unfolded in solution.[10] In intact PSI, these proteins are essential for the final steps of light-induced electron transfer—PsaC coordinates two

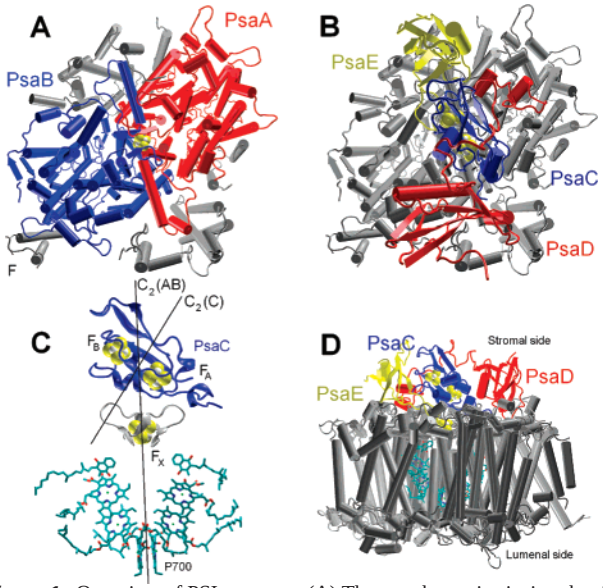

Figure 1. Overview of PSI structure. (A) The membrane-intrinsic subunits of PSI. View is perpendicular to the membrane plane from the stromal side. The C_2 pseudo-symmetry axis runs perpendicular to the page, approximately through the iron−sulfur cluster F_X (shown in yellow and white). (B) Same view as in A, except that the stromal subunits C, D, and E have been added. The C_2 pseudo-symmetry axis is absent in these subunits. (C) Side view of the electron-transfer chain. The C_2 pseudo-symmetry of subunits A and B is exhibited very clearly in the initial cofactors in the electron-transfer chain. Charge separation begins at the special pair P700, and the electron travels up one branch of the electron-transfer chain until it arrives at the iron−sulfur cluster F_X, which is coordinated by loops from PsaA and PsaB (shown in gray). These loops also provide the major interactions with PsaC, which coordinates the distal iron−sulfur clusters F_A and F_B. Note that the area of PsaC that coordinates F_A and F_B exhibits a local C_2 symmetry, with an axis that is at a 62° angle to C_2(AB). (D) Side view of the PSI monomer, from within the membrane plane. Figures 1, 4, 5, 6, 7, and 8 were produced using Raster3D.[2]

[†] Department of Physics & Astronomy.
[‡] Department of Chemistry & Biochemistry.
[§] Center for the Study of Early Events in Photosynthesis.
[¶] Center for Biological Physics.

(1) Wells, S.; Menor, S.; Hespenheide, B.; Thorpe, M. F. *Phys. Biol.* **2005**, *2*, S127−S136.
(2) Merritt, E. A.; Bacon, D. J. *Macromol. Crystallogr., Part B* **1997**, *277*, 505−524.
(3) Fromme, P.; Jordan, P.; Krauss, N. *Biochim. Biophys. Acta* **2001**, *1507*, 5−31.
(4) Jordan, P.; Fromme, P.; Witt, H. T.; Klukas, O.; Saenger, W.; Krauss, N. *Nature* **2001**, *411*, 909−917.
(5) Antonkine, M. L.; Jordan, P.; Fromme, P.; Krauss, N.; Golbeck, J. H.; Stehlik, D. *J. Mol. Biol.* **2003**, *327*, 671−697.

[4Fe-4S] clusters (known as F_A and F_B) which form the terminus of the electron-transfer chain, and the stromal hump is essential for the correct docking of ferredoxin onto PSI.

Biochemical evidence indicates that PsaC, PsaD, and PsaE dock onto the stromal surface of membrane-intrinsic PSI in a well-defined order: C first, followed by D, and then E.[11−14] Magnetic resonance studies[15,16] indicate that the environment of F_A and F_B changes at least three times during assembly: the two clusters are magnetically equivalent for PsaC in solution; they become inequivalent when PsaC is bound onto the membrane-intrinsic portion of the complex, and they assume the magnetic properties of the intact complex only after the binding of PsaD. From this, we can conclude that PsaC undergoes conformational changes which affect the environment of the [4Fe-4S] clusters when it docks onto the membrane-intrinsic portion of PSI, and further conformational changes are induced by the binding of PsaD. The fact that PsaC can be overexpressed in *E. coli* and combined with the membrane-intrinsic portion of PSI in vitro to reassemble functional PSI complexes[17−19] effectively rules out any essential role played by accessory proteins in the association of PsaC with PSI.

The membrane-intrinsic portion of PSI is dominated by its two largest subunits, PsaA and PsaB. These coordinate the earlier elements of the electron transport chain, along with most of the PSI antenna chlorophylls. PsaA and PsaB show extensive similarities in sequence and structure and form a heterodimer with a C_2 pseudo-symmetry axis. The docking site of PsaC lies directly above this C_2 axis and interacts with a group of charged residues on the stromal surface in which this C_2 symmetry is nearly perfect.[5] PsaC, in contrast, shows local C_2 pseudo-symmetry about an axis which is oriented at a 62° angle to the C_2 axis in PsaA/B,[20] and its interaction with its symmetric docking site is highly asymmetric.

The NMR structure of PsaC gives evidence of significant conformational changes between the solution and docked structures.[6] To a first approximation, PsaC in solution can be thought of as a mostly rigid body which coordinates the two iron−sulfur clusters and a flexible C-terminal tail. An extensive analysis of hydrogen-bond contacts between PsaC and PsaA/B in the PSI crystal structure[5] has shown that the only symmetry-breaking element in the interactions between PsaA/B and PsaC is the C-terminal tail, which interacts with a highly specific binding pocket on the surface of PsaB. The symmetry-related site on PsaA contains different residues and does not form the same specific interactions with the C-terminus of PsaC.

After reviewing these data, Antonkine et al.[5] suggested two possible pathways for the docking of PsaC onto PsaA/B.

Body-First Docking: Positively charged residues on the body of PsaC are attracted to negatively charged residues on the stromal surface of PsaA/B. Because PsaC can dock in one of two symmetry-related orientations, both of these docked structures exist in dynamic equilibrium with the unbound PsaC—a given molecule of PsaC is free to bind, dissociate, and bind again with the opposite orientation. The symmetry-breaking tail of PsaC reaches its final position after these contacts have formed, stabilizing the correct orientation and shifting the equilibrium toward it. Further stabilization of the correct orientation of PsaC may occur with the subsequent binding of PsaD, which would serve to lock the correct orientation into place.

Tail-First Docking: The flexible tail of PsaC is the first component to associate with PsaA/B, and symmetry is broken from the very beginning. After the C-terminus has found its binding pocket on PsaB, the body of PsaC comes into contact with the stromal surface of PsaA/B, forming strong hydrogen bonds between the charged residues mentioned above.

Neither of these scenarios is entirely unproblematic. In the body-first scenario, the 12 or more salt bridges between PsaC and PsaA/B may be too strong to allow for a true dynamic equilibrium between the bound and unbound forms of PsaC. In this case, the equilibrium would be shifted so far toward a bound configuration (whether the orientation is correct or not) that an incorrectly oriented PsaC would be unlikely to dissociate and turn around. The problem with the tail-first scenario is exactly the opposite—there is considerable doubt whether the specific interactions between PsaB and the C-terminus of PsaC are strong enough to provide a driving force for association, especially one that can compete kinetically with the electrostatic association. Hydrophobic interactions would certainly play a role in the association of the C-terminus with its binding pocket, but the strength of these interactions is difficult to estimate.

In this study, our goal is to provide further insight into the details of these docking pathways. To this end, we have simulated the docking process using the recently developed constrained geometric simulation algorithm FRODA[1] (Framework Rigidity Optimized Dynamical Algorithm), which is implemented as a module of the FIRST software package (http://flexweb.asu.edu/software/first/). While FRODA does not give information on the energetics of the docking process, it does give us a sense for the conformational space traversed during docking and which molecular events are essential for the successful docking of PsaC.

Theory

FRODA[1] is a Monte Carlo-type algorithm in which diffusive motion in proteins is simulated through random motion of rigid clusters within the protein. The rigid clusters are first identified using a program called FIRST, which identifies rigid and flexible regions in the protein using a graph-theoretical algorithm known as the pebble game. In FIRST, the atoms in the protein structure are used to form the nodes in a network, while the edges between these nodes are formed by covalent and noncovalent

(6) Antonkine, M. L.; Liu, G.; Bentrop, D.; Bryant, D. A.; Bertini, I.; Luchinat, C.; Golbeck, J. H.; Stehlik, D. *J. Biol. Inorg. Chem.* **2002**, *7*, 461−472.
(7) Falzone, C. J.; Kao, Y. H.; Zhao, J.; Bryant, D. A.; Lecomte, J. T. *Biochemistry* **1994**, *33*, 6052−6062.
(8) Barth, P.; Savarin, P.; Gilquin, B.; Lagoutte, B.; Ochsenbein, F. *Biochemistry* **2002**, *41*, 13902−13914.
(9) Mayer, K. L.; Shen, G.; Bryant, D. A.; Lecomte, J. T.; Falzone, C. J. *Biochemistry* **1999**, *38*, 13736−13746.
(10) Xia, Z.; Broadhurst, R. W.; Laue, E. D.; Bryant, D. A.; Golbeck, J. H.; Bendall, D. S. *Eur. J. Biochem.* **1998**, *255*, 309−316.
(11) Li, N.; Zhao, J. D.; Warren, P. V.; Warden, J. T.; Bryant, D. A.; Golbeck, J. H. *Biochemistry* **1991**, *30*, 7863−7872.
(12) Yu, J.; Smart, L. B.; Jung, Y. S.; Golbeck, J.; McIntosh, L. *Plant Mol. Biol.* **1995**, *29*, 331−342.
(13) Chitnis, V. P.; Jungs, Y. S.; Albee, L.; Golbeck, J. H.; Chitnis, P. R. *J. Biol. Chem.* **1996**, *271*, 11772−11780.
(14) Zhao, J.; Snyder, W. B.; Muhlenhoff, U.; Rhiel, E.; Warren, P. V.; Golbeck, J. H.; Bryant, D. A. *Mol. Microbiol.* **1993**, *9*, 183−194.
(15) Vassiliev, I. R.; Antonkine, M. L.; Golbeck, J. H. *Biochim. Biophys. Acta* **2001**, *1507*, 139−160.
(16) Golbeck, J. H. *Photosynth. Res.* **1999**, *61*, 107−144.
(17) Naver, H.; Scott, M. P.; Andersen, B.; Moller, B. L.; Scheller, H. V. *Physiol. Plant.* **1995**, *95*, 19−26.
(18) Fischer, N.; Setif, P.; Rochaix, J. D. *J. Biol. Chem.* **1999**, *274*, 23333−23340.
(19) Rodday, S. M.; Do, L. T.; Chynwat, V.; Frank, H. A.; Biggins, J. *Biochemistry* **1996**, *35*, 11832−11838.
(20) Schubert, W. D.; Klukas, O.; Krauss, N.; Saenger, W.; Fromme, P.; Witt, H. T. *J. Mol. Biol.* **1997**, *272*, 741−769.

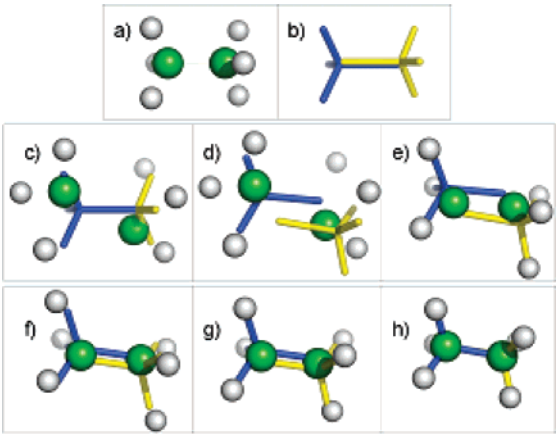

Figure 2. The motion of an ethane molecule as simulated by FRODA. (a) Initial atomic positions; (b) ghost templates; (c) random atomic displacement; (d) fitting of ghost templates to atoms; (e) refitting of atoms to ghost templates; (f) and (g) further iterations of (d) and (e); (h) until a valid new conformer is found. Figure reproduced from Reference 1.

bonds in the protein structure. An important part of the FIRST algorithm, therefore, is the identification of noncovalent interactions—hydrogen bonds and hydrophobic tethers. Once the network has been constructed, the pebble game algorithm is used to identify rigid regions within this network. A detailed description of FIRST and the pebble game approach can be found in refs 21−26.

After clusters of mutually rigid atoms in the protein have been identified by FIRST, they are replaced by ghost templates, which serve to guide the motion of the atoms in the protein (Figure 2A,B). At each step of the simulation, each atom is "thrown" in a random direction, typically over a distance of 0.05−0.2 Å (Figure 2C). Next, the ghost templates are moved to fit the atoms which belong to them (Figure 2D), and the atoms are then moved to their appropriate locations on the ghost templates (Figure 2E). When two ghost templates are joined by a rotatable covalent bond, the atoms participating in this bond will be associated with both ghosts. The sites involved in the shared bond often will not overlap perfectly after the first fitting, so ghosts and atoms are fit iteratively (Figure 2F,G), until atoms and ghosts overlap each other within a certain acceptable tolerance, usually 0.125 Å (Figure 2H). The end result is that rigid units within the protein move randomly by rotation about covalent bonds.

Two targeting schemes—direct targeting and simulated annealing—allow users to direct the motion of a protein toward a target conformer. In direct targeting, a small bias is superimposed upon the random throws for each atom. Each atom in the initial structure is biased toward its location in a target structure, causing the system to drift toward its final configuration. Random motion is still present, which allows the system to make conformational changes more complex than could be achieved by straight-line motion of atoms. In simulated annealing, the atom throws are completely random, but a given protein conformer, once generated, is accepted or rejected according to a criterion similar to the Metropolis criterion,[27] except that the relevant quantity is the root-mean-squared deviation from the target, rather than the energy. In other words, a random move which moves the system closer to its target structure (i.e., the RMSD-to-target value decreases, $\Delta_{RMSD} < 0$) will always be accepted, and a random move which moves the system away from the target ($\Delta_{RMSD} > 0$) will be accepted with a probability of $e^{-\Delta RMSD/l}$, where the annealing length scale l plays a role similar to that played by the thermal energy kT in the Metropolis algorithm. Typical values of l range from 0.001 to 0.02 Å, while typical values for Δ_{RMSD} are on the order of 0.001−0.01 Å.

A distinction should be made between the simulated annealing approach used by FRODA and the standard Monte Carlo approach which uses an energy functional in the evaluation of the Metropolis criterion.[28] The latter involves rugged energy landscapes[29] with many basins and barriers, and parameter sensitivity can be rather high. In FRODA, however, our annealing takes place on a simple RMSD-to-target. This gives a much simpler and less textured landscape and, correspondingly, lower parameter sensitivity. When simulated annealing steps, such as those described in this paper, were repeated with a variety of parameters, we found that a broad range of parameters (a random step size of 0.01−0.1 Å and an annealing length scale l of 0.001−0.015 Å) was able to produce positive results. For a simulation with a given set of parameters, the main determinant of the success or failure of the targeting procedure was the random seed used—simulations with identical parameters could either converge to their target or develop irresolvable steric problems, based on which random steps are taken. Successful simulations with different random seeds will vary in the time needed to converge and the details of the final structure, but show qualitatively similar behavior.

It is helpful to compare the FRODA approach to the more traditional approach, molecular dynamics, which is based on the evaluation of a many-body energy function for each conformation. The traditional approach allows the calculation of a free energy landscape with numerous peaks and valleys; when the system undergoes a conformational change, it moves from one local minimum to another along a path that minimizes the increase in free energy needed to move from one minimum to another[29,30] (Figure 3A). One of FRODA's main advantages is that it is dramatically faster than molecular dynamics, enabling the exploration of large-amplitude motions characteristic of longer time scales with larger systems than is possible with molecular dynamics. This increased speed is a result of FRODA's simplified energy function; rather than evaluating numerous nonbonded interactions for each atom in the system, conformers are simply classified as allowed or nonallowed based on whether they can be reached by random throws that keep all atoms on their ghost templates. FRODA does not give us any further information on the energy of different conformational states, but efforts are underway to utilize force-field potentials in future implementations of FRODA. At present, allowed conformers

(21) Jacobs, D. J.; Thorpe, M. F. *Phys. Rev. Lett.* **1995**, *75*, 4051−4054.
(22) Jacobs, D. J.; Thorpe, M. F. *Phys. Rev. E* **1996**, *53*, 3682−3693.
(23) Jacobs, D. J.; Rader, A. J.; Kuhn, L. A.; Thorpe, M. F. *Proteins* **2001**, *44*, 150−165.
(24) Rader, A. J.; Hespenheide, B. M.; Kuhn, L. A.; Thorpe, M. F. *Proc. Natl. Acad Sci. U.S.A.* **2002**, *99*, 3540−3545.
(25) Hespenheide, B. M.; Rader, A. J.; Thorpe, M. F.; Kuhn, L. A. *J. Mol. Graph. Model* **2002**, *21*, 195−207.
(26) Hespenheide, B. M.; Jacobs, D. J.; Thorpe, M. F. *J. Phys.: Condens. Mater.* **2004**, *16*, S5055−S5064.
(27) Metropolis *J. Am. Stat. Assoc.* **1949**, *44*, 335.
(28) Press, W. H. *Numerical recipes: the art of scientific computing*; Cambridge University Press: Cambridge, New York, 1986.
(29) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598−1603.
(30) Wales, D. J. *Philos. Trans. A Math Phys. Eng. Sci.* **2005**, *363*, 357−375; discussion 375−357.

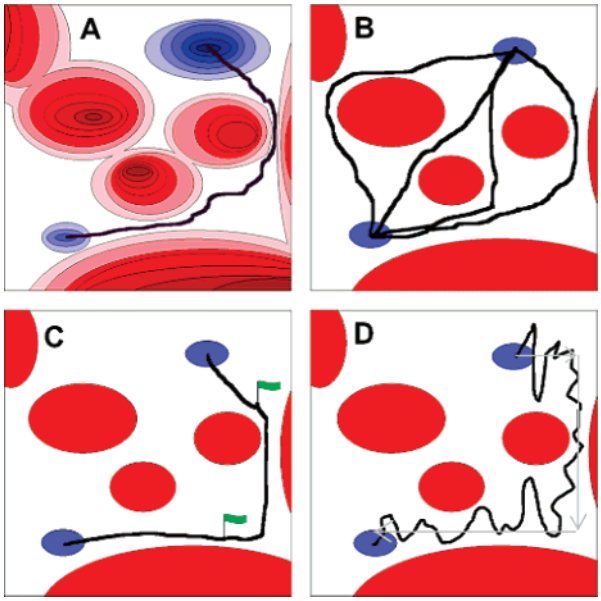

Figure 3. (A) Schematic of a two-dimensional energy landscape. Maxima are shown in red, minima are shown in blue. Black line shows the favored transition pathway from one conformational state to another. (B) Energy landscapes in FRODA. The simplified energy function used in FRODA results in a given point in phase space being allowed (white) or disallowed (red). The loss of information results in a larger number of possible pathways, with no a priori way to distinguish between them. (C) Choosing a pathway by targeting to intermediate structures. The green flags specify completely specified intermediate structures; the system moves to each of these before being targeted to the final state. (D) Choosing a pathway by partial targeting. The system moves from its initial to its final state in three steps. In each step, one dimension is targeted (first horizontal to the right, then vertical downward, then horizontal to the left), while the other fluctuates randomly. This does not require complete knowledge of intermediate structures.

effectively have an energy of zero, and nonallowed ones have infinite energy.

This computational speedup, however, comes at a cost. Because FRODA uses such a simple energy function, a detailed knowledge of the free energy landscape cannot be obtained. This means that, out of several possible pathways between two points on the free energy landscape, a quantitative distinction in terms of the height of the free energy barrier to be crossed cannot be made. FRODA can only make the qualitative distinction that a given conformer is allowed or disallowed; hence a given pathway can only be labeled as possible or impossible. The biologically relevant pathway will certainly be a possible one, but some possible pathways may turn out to be biologically irrelevant because they require the crossing of a free energy barrier that FRODA did not notice (Figure 3B).

If we have some knowledge of intermediate states in the transition, then we can cause FRODA to choose one of these possible paths. This does not give us a conclusive test of whether a proposed pathway is biologically relevant, but it does allow us to examine individual hypothetical pathways and learn more about their molecular details. It may also allow us to rule out some potential pathways as being sterically disallowed. One method would be to specify an intermediate point in conformation space and target our system to this intermediate structure before going to the final one (Figure 3C). For example, for a protein with open and closed forms, we would target the protein from its open state to a halfway-open state and finally to the closed state. The major drawback to this method is that it requires knowledge of intermediate conformations of the protein, which is often unavailable.

Another approach, rather than specifying intermediate states of the system, is to apply targeting to only part of the system (Figure 3D). In effect, the targeting takes place on a lower-dimensional subspace of the free energy landscape. In practice, this involves targeting only some of the atoms in the system to specific locations. Those atoms not directly involved in the targeting will simply respond as dictated by their bonding relationships. By targeting only a small number of residues in a protein, this method allows us to see which residues can be used to steer conformational changes and which residues are more passive participants in the change.

Methods

All simulations were performed using the command-line version of FIRST 5.2, available at http://flexweb.asu.edu/software/first. Each phase described above took roughly 1−20 h on a single Linux processor. In both docking scenarios (tail-first vs body-first), the first step was performed using simulated annealing (as described above), while the subsequent steps relied on direct targeting of the affected residues.

After noncovalent bonding interactions (e.g., hydrogen bonds, hydrophobic tethers) have been identified by FIRST, the same bond network is maintained throughout the course of a FRODA simulation.[1] If applied naïvely, this can lead to an inaccurate picture of protein conformational changes since these weaker interactions can be broken and re-formed in the course of a conformational change.[31] The usual method applied when using FIRST/FRODA is to identify the noncovalent bond networks of the initial and final structures and to utilize a bond network containing only those constraints held common by the two; this allows for the sort of "local unfolding" that often takes place during large conformational changes.

In the present study, a more nuanced approach was needed. First of all, the method of keeping the same set of noncovalent bonds throughout the docking process is linked to the direct-targeting method of simulating conformational changes, in that it requires knowledge of the structure of the final state which is not necessarily present when performing partial targeting. Another consideration for the docking of PsaC, in particular, is that, although the NMR and bound crystal structures are qualitatively similar, the overlap between their noncovalent bond networks is very small. Initial FRODA simulations showed that the set of shared noncovalent bonds was not enough to keep the protein folded. One reason for this is the large conformational changes between the NMR and crystal structures, especially in the N-terminus, which forms part of the hydrophobic core in the NMR structure and is located on the surface of the protein in the crystal structure. Another reason is that the stereochemical quality of the crystal data is generally higher than the NMR data—it is possible that the quality of the NMR models is not sufficient to accurately identify the hydrogen bonds present in solution.

We therefore used an approach in which the noncovalent bond network was somewhat different during each stage of the simulations, as would be expected for a protein undergoing complex conformational changes. In both docking scenarios, the first phase involved the initial docking of PsaC onto PsaA/B. We assumed that the conformational changes in PsaC take place because of its association with PsaA/B, rather than taking place spontaneously in solution before docking. As a result, noncovalent constraint parameters were chosen which caused FIRST to identify the "body" of PsaC, which binds the iron−sulfur clusters, as a single rigid unit to which the flexible C-terminus is attached. More specifically, when FIRST was run with a hydrogen bond cutoff of −1.0 kcal/mol (i.e., all hydrogen bonds weaker than this were ignored) and noncovalent interactions were identified by a procedure

(31) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12570−12575.

Table 1. Details of Targeting and Bond Network Evolution during Each Step of the Docking Simulations (the numbers in parentheses indicate the number of noncovalent contacts conserved from the first phase of the simulation)

		Phase 1	Phase 2	Phase 3	Phase 4
Tail-first docking	Residues targeted	Tyr 80 targeted to final position	Lys 51 and Arg 52 targeted to final positions	Ala 70 through Tyr 80 targeted to final positions	Ala 1 through Tyr 7, Leu 25, Arg 65 targeted to final positions
	Intra PsaC	7 H-bonds 98 Hydrophobic	10 H-bonds (7) 29 Hydrophobic (19)	13 H-bonds (6) 30 Hydrophobic (19)	12 H-bonds (6) 27 Hydrophobic (18)
	PsaC-core	0 H-bonds 0 Hydrophobic	1 H-bond 3 Hydrophobic	6 H-bonds 6 Hydrophobic	5 H-bonds 9 Hydrophobic
	Core	1369 H-bonds 2120 Hydrophobic	1386 H-bonds (1368) 2102 Hydrophobic (805)	1393 H-bonds (1368) 2121 Hydrophobic (762)	1406 H-bonds (1364) 2256 Hydrophobic (750)
Body-first docking	Residues targeted	Lys 51 and Arg 52	Glu 54 and Arg 65	Ala 1 through Tyr 7, Leu 25, Ile 64 through Ala 70	Ala 70 through Tyr 80
	Intra PsaC	7 H-bonds 95 Hydrophobic	10 H-bonds (7) 37 Hydrophobic (20)	10 H-bonds (6) 22 Hydrophobic (14)	14 H-bonds (6) 37 Hydrophobic (11)
	PsaC-core	0 H-bonds 0 Hydrophobic	4 H-bonds 0 Hydrophobic	15 H-bonds 1 Hydrophobic	11 H-bonds 3 Hydrophobic
	Core	1369 H-bonds 2120 Hydrophobic	1410 H-bonds (1366) 2238 Hydrophobic (792)	1416 H-bonds (1367) 2313 Hydrophobic (786)	1420 H-bonds (1366) 2363 Hydrophobic (782)

similar to that used previously,[25] a rigid core was identified that contained 56% of the atoms in residues 1−68 and only 1.2% of the atoms in residues 69−80. This helped avoid any internal conformational changes in the body of PsaC until interactions with PsaA/B had actually begun.

For the subsequent phases of the docking simulations, the following procedure was used. First, the starting structure for the phase being simulated (i.e., the finishing structure of the previous phase) was input into FIRST and used to generate a list of noncovalent constraints. Second, those constraints which involved residues which were actively targeted in the phase being simulated were removed. This allowed the residues being actively targeted to move freely and reach their target locations. Any cooperative motion of the nontargeted residues was therefore mediated by their covalent connections to the residues being actively targeted. Once the residues being actively targeted had reached their intended positions, this final structure was used to calculate the noncovalent bond network for the next phase of the simulation. By using this method, the noncovalent bond network was able to change in a piecewise manner—after each phase of the simulation, the noncovalent bond network had been modified in a way that reflected the conformational changes which had occurred during that phase. Most of the partial-targeting phases were intended to form specific noncovalent interactions, so this sort of continually evolving bond network was helpful in making these changes permanent.

The data shown in Table 1 offer us some important insights into the evolution of the bond network. Because the first step in both scenarios involved large-scale translational motion of PsaC rather than detailed conformational changes, hydrophobic tethers were chosen in a way that made most of PsaC a single rigid unit, as mentioned above. In subsequent steps, more restrictive criteria were used to identify hydrophobic tethers; this led to a decrease in the number of hydrophobic tethers included, especially within PsaC. One thing that is immediately apparent is the net increase in the number of hydrogen bonds and hydrophobic tethers during the course of the simulations. This increase is apparent even in PsaA/B, where no residues were targeted and all motion present was unbiased. This can be explained by noting that our method of bond network evolution allows new noncovalent bonds to form when previously unconnected residues come close enough together that they are bonded at the beginning of the next phase of the simulation. Only rarely will the geometry of bonded atoms change enough during one phase of the docking that that bond will be absent in the next phase. FRODA is rather strict about maintaining the geometry of hydrogen bonds, so in each case, new hydrogen bonds are added to the network, while most of the initial set is maintained. The geometric constraints on hydrophobic tethers are less strict, so it is easier for hydrophobic tethers to be broken when the bond network is re-evaluated at the beginning of each new phase. Although no significant conformational changes occur in PsaA/B, the total number of hydrophobic tethers increases at each step, while the fraction of the original set of tethers maintained in subsequent steps decreases. This increase in the number of noncovalent interactions will lead to an increase in the overall rigidity of the protein; in our case, this turned out not to be a problem since the docking site on PsaA/B was largely rigid to begin with.

Results

Rigidity Analysis. After hydrogen atoms had been added to the PSI crystal structure using Reduce,[32] a FIRST analysis with default settings (including hydrogen bonds stronger than −1.0 kcal/mol) indicated that PSI is dominated by a single rigid unit that includes 58% of the atoms in the structure, including 70% of the atoms in PsaA and 75% of PsaB. This is consistent with what is known about the function of PsaA and PsaB; their role in light-harvesting does not involve conformational changes, and their structures are largely rigid, without any moving parts. PsaC participates in this rigid cluster to a lesser extent—only 48.8% of its atoms are mutually rigid with the bulk of the complex.

When the membrane-extrinsic subunits PsaD and PsaE are removed from the structure, 70% of PsaA and 75% of PsaB remain involved in the largest rigid unit, while only 12% of PsaC stays rigid. This is consistent with our expectation that PsaD and PsaE play an important role in the stabilization of PsaC and have relatively little impact on the rigidity of PsaA and PsaB. Because PsaC docks onto the complex before PsaD and PsaE, it is this partial structure, lacking PsaD and PsaE, that will be relevant for our docking simulations. It is noteworthy that, without the stabilizing influence of PsaD and PsaE, the peptide backbone of PsaC participates in the rigid core in only two sections: the α-helical region from Lys51 to Ala56 and the β-sheet region from Arg65 to Tyr67. Both of these regions will prove to be important in the dynamical docking simulations, as they contain charged residues which form strong salt bridge interactions with PsaA/B.

In the hypothesized docking mechanisms of PsaC, the flexible C-terminus (beginning roughly at Leu68) is thought to play an important role as a symmetry-breaking element. It is interesting to note that, despite the crucial structural role of this C-terminus,

(32) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735−1747.

only one group of atoms participates in the main rigid core: the phenol group of Tyr80. While FIRST identifies only four noncovalent bonds between residues 68−79 and their binding site on the surface of PsaB, six noncovalent bonds are identified between Tyr80 and the binding site. More specifically, two hydrophobic contacts are identified with Val685 and one with Glu682 of PsaB, and strong hydrogen bonds are formed between Tyr80 and Lys702 (−1.7 kcal/mol), Gln678 (−3.4 kcal/mol), and Pro703 (−6.0 kcal/mol) of PsaB.

In the rigidity analysis of the complete structure, including PsaD and PsaE, the backbone of PsaC is flexible in only three regions: the loop regions from Asp8 to Pro21, from Val28 to Ser40, and from Cys57 to Ile64. Almost all of the atoms in the C-terminus (87%) are part of the main rigid cluster. These differences in rigidity indicate that Tyr80 may play an especially significant role in the early steps of the docking process, but that most of the C-terminus remains flexible until PsaD has docked. The docking of PsaD leads to the rigidification of much of PsaC, in particular, the C-terminal region.

These insights into the important structural role of Tyr80 lead to a more biologically oriented question: How well-conserved is Tyr80 among PsaC sequences from different species? A ClustalW[33] comparison of complete or fragmentary PsaC sequences identified using BLAST[34] (115 total) indicated that Tyr80 is conserved among all but one species for which the C-terminus of PsaC has been sequenced; the complete multiple alignment along with a list of pairwise alignment scores is included in the Supporting Information for this article. While this degree of conservation is impressive, it must be kept in mind that PsaC is a generally well-conserved protein; an alignment of complete *psaC* sequences (75 total) shows that 54% of residues in the protein are universally conserved. These 75 sequences came from a diverse set of organisms, including numerous species of plants, cyanobacteria, red and green algae, and marine diatoms.

Dynamics. We have investigated the two proposed scenarios for the docking of PsaC using the constrained geometric simulation algorithm FRODA. Structures are available for PsaC in solution (PDB ID 1K0T) and PsaC as part of the fully assembled PSI complex (PDB ID 1JB0), but no structures are available for "halfway-docked" PsaC. As a result, the best possibility for investigating details of the PsaC docking pathway is using the partial targeting approach described in the Theory section above.

As a preliminary step, we performed a direct docking of PsaC onto PsaA/B. The initial structure included PsaA and PsaB (from 1JB0) and one of the NMR models of the PsaC solution structure, located about 35 Å above its stromal docking site. The system was simulated with all atoms in PsaC biased toward their locations in the docked crystal structure. PsaC was able to successfully reach its target conformation (Figure 4). The pathway obtained by this process did not resemble either of the proposed docking mechanisms—the entire protein arrived more or less simultaneously, rather than in identifiable steps—but the direct-targeted simulation was able to prove that at least one possible docking pathway exists, which is a prerequisite to being able to identify the biologically relevant pathway. Once this was established, we proceeded to examine the two proposed docking mechanisms in detail. See Table 1 for a summary of the steps in these two mechanisms.

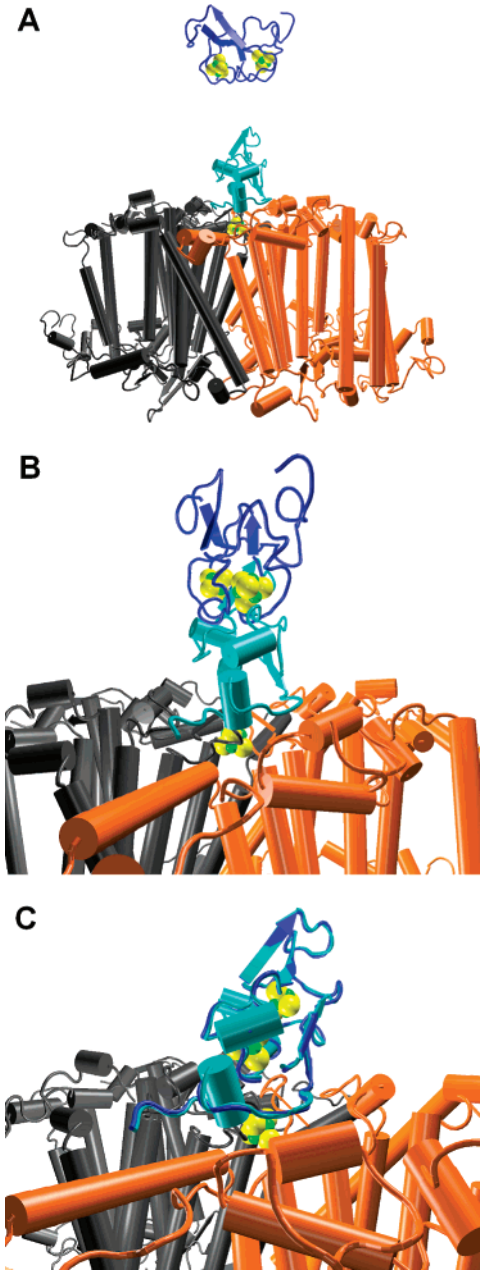

Figure 4. Direct docking of PsaC onto PsaA/B. Each atom in PsaC (blue) was biased toward its target location (cyan) at each step in the simulation. The starting structure for PsaC was one of the solution NMR conformers, while the final structure was the crystal structure bound to PsaA/B. An animated version of this figure is available in the Supporting Information accompanying this article as Movie S1.

Tail-First Docking. The first step in the tail-first docking was to target Tyr80, the C-terminal residue of PsaC, to its binding site on the stromal surface of PsaB using simulated annealing (Figure 5A,B). Tyr80 was chosen as the residue to initiate docking for the reasons mentioned in the Rigidity Analysis section above. The targeting of PsaC to its final binding site led to an extension of the C-terminus, giving the appearance that PsaC was "harpooning" the larger membrane proteins.

In the next step of the tail-first docking, the positively charged residues Lys51 and Arg52 were targeted to the positions that

(33) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.
(34) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.

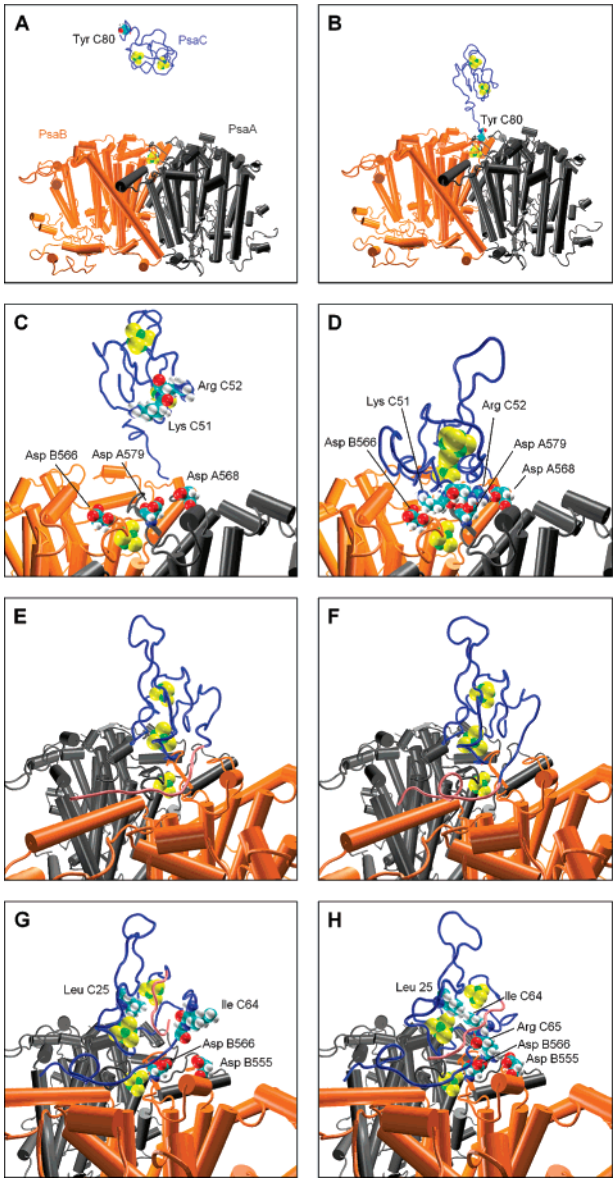

Figure 5. Tail-first docking of PsaC. (A) Initial state, the same as in Figure 3. (A, B) C-terminal residue Tyr80 on PsaC is targeted to its binding site on PsaB. (C, D) Lys51 and Arg52 are targeted to their final position, where they interact with charged residues on PsaA/B. (E, F) Optimization of contacts between the C-terminus (pink) of PsaC and PsaB. (G, H) The N-terminus of PsaC (pink) leaves its internal location and is targeted toward its final location on the surface of PsaC. Animated images of the transitions between the pairs of images mentioned above are available as Movies S2a (for A, B), S2b (for C, D), S2c (for E, F), and S2d (for G, H) in the Supporting Information accompanying this article.

they occupy in the PSI crystal structure (Figure 5C,D). These form many of the important salt bridges between PsaC and PsaA/B; in particular, they interact with Asp568 and Asp579 on PsaA and Asp566 on PsaB. A thorough examination of the strong hydrogen bond interactions between PsaC and PsaA/B[5] indicates that Lys51 and Arg52 form one side of the basic binding patch on the underside of PsaC, with the remainder of this patch being formed by Arg65.

The crucial role played by these charged residues is underscored by an intriguing mutagenesis study,[18] in which two of them (Lys52 and Arg53) were replaced by neutral ones in a V49I/K52T/R53Q triple mutant in *Chlamydomonas reinhardtii*. The resulting mutant organisms were unable to grow photoautotrophically and were photosensitive at high light intensities, indicating a defect in photosynthetic activity. The triple mutant also accumulated subunits PsaD and PsaE in much lower levels than in the wild type, indicating that the mutation destabilized the stromal hump. PSI isolated from the triple mutant showed a 3-fold decrease in its affinity for ferredoxin, giving further evidence for the importance of these residues in the assembly of the stromal hump.

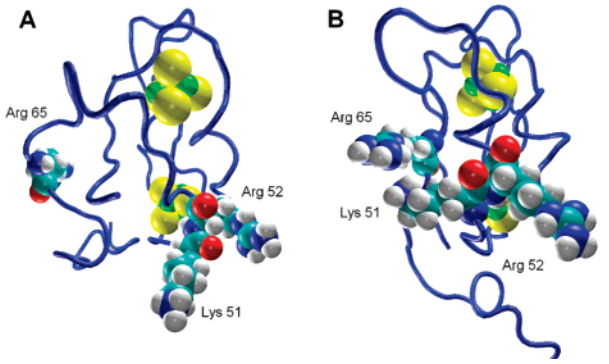

Figure 6. Charged residues on PsaC. The positively charged residues Lys51, Arg52, and Arg65 on PsaC form the strongest salt bridge interactions between PsaC and PsaA/B. They are oriented differently in the NMR (A) and X-ray (B) structures.

Comparison of the docked crystal structure of PsaC and one of the solution NMR structures reveals that the spatial relationship of these residues is rather different in the two structures (Figure 6). In the final docked structure, these residues are placed rather close together and interact with closely spaced negatively charged residues on the stromal surface of PsaA/B. In the solution structure, Arg65 is located about 19 Å from Lys51 and Arg52, with the space between them being taken up by the N-terminus of PsaC, which is rearranged in the docked structure to form an antiparallel β-sheet with part of the C-terminus.

Some more changes therefore needed to take place before Arg65 could move into position, completing the formation of the strongest interactions between PsaC and PsaA/B. In the third step of the tail-first docking (Figure 5E,F), the C-terminal residues Ala70 through Tyr80 were targeted to their final positions. This served to optimize the contacts between PsaC and PsaB, as well as position residues Arg65 through Ala70 to form a β-sheet with the N-terminus of PsaC. During the fourth and final step (Figure 5 G,H), these residues, along with the N-terminal residues Ala1 through Tyr7, were targeted to their final positions. After the N-terminus of PsaC was moved from its internal position to one in which it is located on the surface of the protein, Arg65 was moved to its final position, in which it forms salt bridges with Asp555 and Asp566 of PsaB.

One subtle feature of the conformational changes in PsaC appeared during the final step of the PsaC docking. When only the residues mentioned above were targeted to their final positions, the rest of the protein basically fell apart—the integrity of PsaC's basic fold was not preserved. The reason for this was that in the NMR structure the N-terminus is buried in the hydrophobic core of the protein and thus forms many of the hydrophobic interactions that help PsaC maintain its structural integrity. When the N-terminus is removed, the integrity of the hydrophobic core is lost, and the fold comes apart. We found that, in the crystal structure, FIRST identified hydrophobic interactions between Leu25 and Ile64, which are absent in the

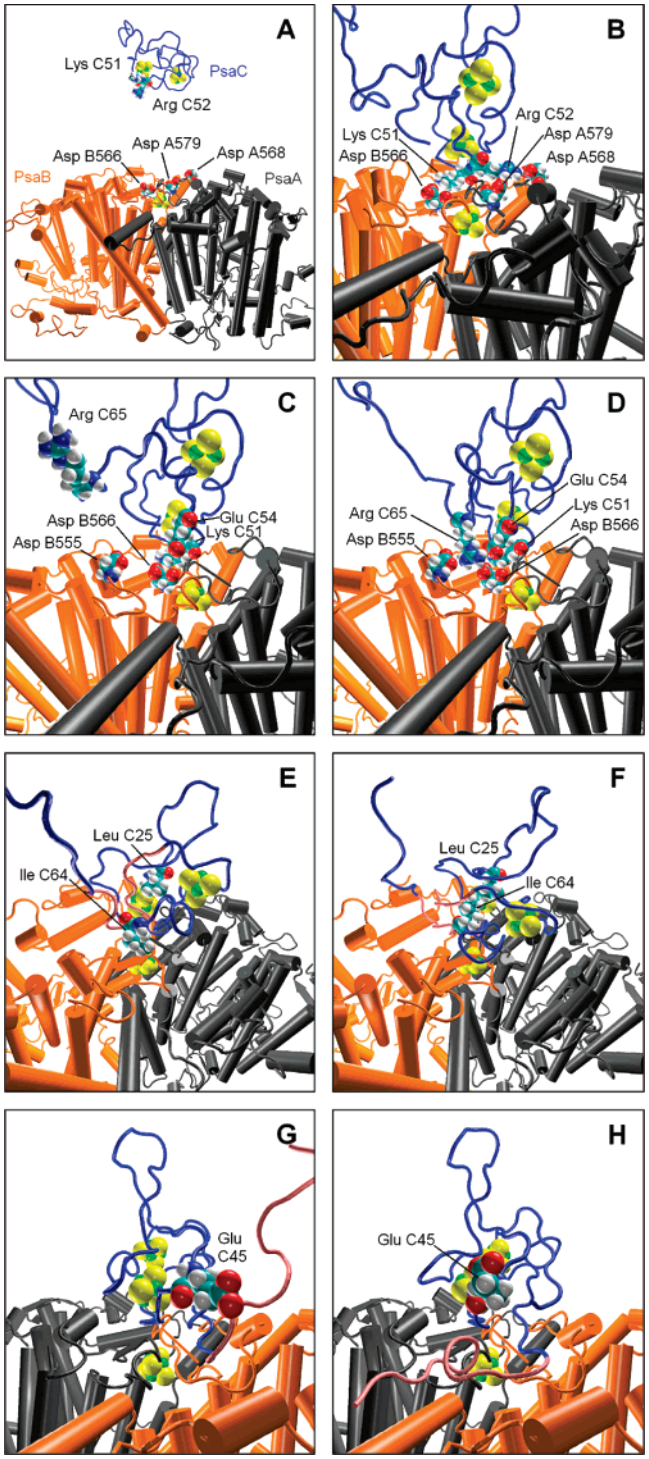

Figure 7. Body-first docking of PsaC. (A, B) Lys51 and Arg52 targeted to their final positions. (C, D) Glu54 and Arg65 targeted to their final positions. (E, F) N-terminus of PsaC (pink) is moved to its final position. (G, H) C-terminus of PsaC (pink) moves to its binding site on PsaB. Animated images of the transitions between the pairs of images mentioned above are available as Movies S2a (for A, B), S2b (for C, D), S2c (for E, F), and S2d (for G, H) in the Supporting Information accompanying this article.

solution NMR structure. If Leu25 is also targeted to its final position during the last step of the simulation, then new hydrophobic bonds are formed to replace those which are disrupted, and the protein fold remains intact.

Body-First Docking. In the other scenario for the docking of PsaC, the strong salt bridges between PsaC and the stromal surface of PsaA/B form first, and the exact positioning of the flexible tail is the final step. The first step of the body-first docking (Figure 7A,B) is similar to the second step of the tail-first docking—Lys51 and Arg52 are targeted toward their final positions using simulated annealing. Because the flexible C-terminus of PsaC is not being directed anywhere in particular, it tends to trail behind the body of PsaC, which is being pulled forward by these two residues. In the next step (Figure 7C,D), these two residues were held in place while Glu54 and Arg65 were targeted to their final positions. Glu54 was included in this step for reasons similar to those for which Leu25 was included in the fourth step of the tail-first docking; it forms hydrogen bonds with Lys51 that are important to maintaining the integrity of PsaC's fold. There is enough flexibility present in the structure of PsaC that Arg65 could reach its final position without requiring that the N-terminus be moved from its internal position.

In the third step (Figure 7E,F), residues Ala1 through Tyr7, Leu25, and Ile64 through Ala70 were targeted to their final positions. This step was very similar to the final step of the tail-first docking scenario; the N-terminus was moved from its internal position to form a β-sheet on the surface of the protein, and Leu25 and Ile64 were used to maintain the integrity of the hydrophobic core. In the final step (Figure 7G,H), the flexible C-terminus, consisting of residues Ala70 through Tyr80, was targeted to its binding site on the stromal surface of PsaB.

Discussion

In this study, we have demonstrated the usefulness of the constrained geometric simulation algorithm FRODA for the investigation of alternative pathways in a protein-protein docking problem. The amount of information used to outline the two alternative pathways was minimal—for the tail-first docking pathway, we only knew that the flexible C-terminus docks before the main body of the protein, and for the body-first docking pathway, we knew that the charged residues on the main body of the protein form their strong interactions before the flexible C-terminus reaches its binding pocket.

Before any dynamical calculations were performed, however, a static rigidity analysis of the PSI structure using FIRST proved to be very fruitful. This static analysis was performed very quickly (the calculations on the complete PSI crystal structure took about 15 s on a single processor), but yielded a wealth of information that was consistent with experimental results. In particular, we found that PSI is dominated by a single, monolithic rigid unit that involves more than 70% of the atoms in the largest transmembrane subunits, PsaA and PsaB. While only a small fraction of PsaC (11.7%) is part of this major rigid cluster when PsaD and PsaE are absent from the structure, their presence causes more of PsaC to become mutually rigid with PsaA and PsaB (increasing to 48.8%), while PsaA and PsaB are largely unaffected. The difference in the C-terminus of PsaC was particularly dramatic; it is almost completely flexible when PsaD and PsaE are absent and almost completely rigid when they are present.

The static rigidity analysis, together with a multiple alignment of PsaC sequences from a wide variety of species, also pointed to a crucial role of Tyr80 in the docking of PsaC. Tyr80 was the only residue in the C-terminus of PsaC to participate in the major rigid cluster in the absence of PsaD and PsaE, suggesting that it may be uniquely involved in the early steps of PsaC docking. FIRST was also able to identify a much larger number of specific contacts between Tyr80 and its binding site on PsaB

than for the rest of the C-terminus (six for Tyr80 alone, compared to four for the preceding 12 residues), further underscoring its unique role.

By studying partial-targeted simulations of PsaC docking using FRODA, we were able to elucidate details about the docking pathway that were not apparent from simple comparison of the initial and final structures. For example, when the N-terminus of PsaC is moved from its interior position in the solution structure to its surface position in the docked structure, it appears that the formation of hydrophobic interactions between Leu25 and Ile64 is necessary to maintain the integrity of the protein fold. We also found that the requirement that Arg65 move closer to Lys51 and Arg52 during the docking process does not place constraints on the rearrangement of the N-terminus; PsaC is flexible enough that the N-terminus can move either before or after Arg65 reaches its final position. It appears that the really essential distinction between the two scenarios is whether the main body of electrostatic contacts is formed before or after the symmetry breaking by the C-terminus—the timing of the N-terminal rearrangement is not critical.

It is worthwhile to compare the results of these two docking pathways. After the simulation steps described above, the overall backbone RMSD between the PsaC structure obtained by tail-first docking and the crystal structure was 5.17 Å compared to 6.00 Å for the structure obtained by body-first docking. The tail-first scenario comes closer to the crystal structure despite fewer residues being actively targeted: a total of 23 compared to 30 in the body-first case. Since experimental evidence indicates that PsaC does not assume its final structure until the binding of PsaD, no final targeting of these structures to the position of PsaC in the PSI crystal structure was performed. In the structures obtained from both simulations, the region of PsaC that interacts directly with PsaA/B was much more similar to the crystal structure than the loop regions located further away from PsaA/B (Figure 8). Note also that some of the regions of lower agreement correspond roughly to the loop regions of PsaC which were not rigid in the native PSI crystal structure: residues 8−21, 28−40, and 57−64 of PsaC.

One important difference between the results of the two simulations is in the residues that coordinate the distal iron−sulfur complex F_B. F_B is coordinated by C10, C13, C16, and C57 of PsaC, none of which were actively targeted at any step in the simulation. Although the RMSD between the docking results and the crystal structure are similar for the region C10−C16 (Figure 8A), the backbone conformation for the tail-first docking (Figure 8C) is more similar to the crystal structure (Figure 8A) than the result of the body-first docking (Figure 8D). Qualitatively, the backbone conformation for the tail-first docking seems to be more similar to the crystal structure than the conformation obtained through body-first docking.

Although the simulations do not allow us, at this stage, to decide between these two docking pathways, the above considerations lead us to favor the tail-first docking pathway over the body-first docking pathway. A definitive decision in this regard, however, can best be made experimentally. By giving specific details for a hypothetical pathway, the constrained geometric simulations can be very helpful in the design of experiments that could distinguish between alternative pathways. For example, Tyr80 is of absolutely crucial importance in the tail-first docking pathway, but only secondary importance in the body-first docking pathway; a mutagenesis experiment where this residue is deleted could provide insight into its role.

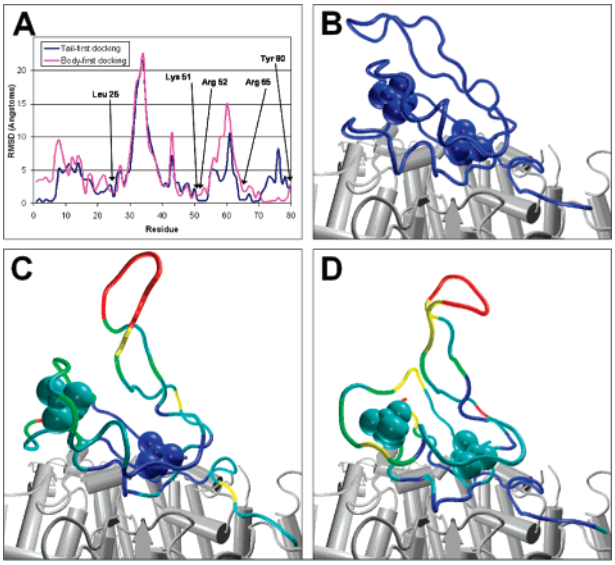

Figure 8. Comparison of the final structures obtained to the docked PsaC crystal structure. (A) Root-mean-square deviation between the final structures obtained in the docking simulations and the PsaC crystal structure. The RMSD in Å is plotted against the residue number, with the tail-first docking in blue and the body-first docking in pink. The labeled residues are, from left to right, Leu25, Lys51, Arg52, Arg65, Tyr80. Note the general agreement between the two simulation methods, where residues which were actively targeted are close to their positions in the crystal structure, while those that were not are in conformations less similar to the crystal structure. Also note that the C-terminus is much closer to its location in the crystal structure in the body-first scenario than in the tail-first scenario; the opposite is true for the N-terminus. In each case, the portion which was targeted last in the simulation is in better agreement. (B) Docked PsaC crystal structure (for comparison). (C) Results of tail-first docking of PsaC. Backbone and iron−sulfur clusters are color-coded according to RMSD from crystal structure (RMSD > 10 Å in red, 7−10 Å in yellow, 5−7 Å in green, 2−5 Å in cyan, and RMSD < 2 Å in blue). (D) Results of body-first docking of PsaC; coloring scheme is the same as that in C.

Similarly, Leu25 is broadly conserved among PsaC sequences from a variety of species, and mutagenesis studies could help to further clarify its role in the conformational changes of PsaC. There are hydrophobic interactions between Val4 (which is universally conserved) and Leu25 in the solution structure, but Leu25 forms hydrophobic interactions with Ile64 (which is sometimes replaced with a Val) in the docked structure. Val4 also forms strong hydrophobic interactions with the universally conserved Leu68 in solution. If Val4 were to be replaced with a less hydrophobic residue, would PsaC adopt something more like its docked structure in solution, or would it fall apart completely? Would replacement of Ile64 by a polar residue prevent PsaC from docking correctly?

The FRODA algorithm is undergoing continued development and may prove to be useful in a variety of related problems in protein folding, complex assembly, and conformational changes. Both a command-line version and an interactive web interface for the FIRST software, which includes FRODA, are available at http://flexweb.asu.edu/software/first. Investigations of the docking of PsaD and PsaE with FIRST/FRODA are currently underway. While the partial-targeting method may not allow us to unequivocally predict the details of a conformational transition, it can be a very powerful tool to explore a hypothetical transition pathway.

Acknowledgment. This work was supported by NSF Grant MCB-0417142 and USDA Grant 2003-35318-13573 to P.F., as well as NSF Grant DMR-0078361 and NIH Grant GM067249 to M.F.T. C.J. was supported by an NSF IGERT fellowship. We would also like to thank Dr. Dietmar Stehlik and Dr. Mikhail Antonkine for helpful discussions and for providing NMR constraint data.

Supporting Information Available: Movies of the direct docking of PsaC (similar to Figure 4), the tail-first docking of PsaC (similar to Figure 5), and the body-first docking of PsaC (similar to Figure 7). Multiple sequence alignment of available PsaC sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0587749