JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Structure of the Neisserial Outer Membrane Protein Opa_{60}: Loop Flexibility Essential to Receptor Recognition and Bacterial Engulfment

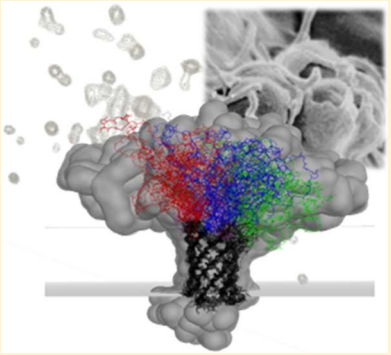
Daniel A. Fox,[†] Per Larsson,[‡] Ryan H. Lo,[†] Brett M. Kroncke,[†] Peter M. Kasson,[‡] and Linda Columbus*[,†,‡]

[†]Department of Chemistry, University of Virginia, Charlottesville, Virginia 22904, United States

[‡]Center for Membrane Biology and Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908, United States

Supporting Information

ABSTRACT: The structure and dynamics of Opa proteins, which we report herein, are responsible for the receptor-mediated engulfment of *Neisseria gonorrheae* or *Neisseria meningitidis* by human cells and can offer deep understanding into the molecular recognition of pathogen−host receptor interactions. Such interactions are vital to understanding bacterial pathogenesis as well as the mechanism of foreign body entry to a human cell, which may provide insights for the development of targeted pharmaceutical delivery systems. The size and dynamics of the extracellular loops of Opa_{60} required a hybrid refinement approach wherein membrane and distance restraints were used to generate an initial NMR structural ensemble, which was then further refined using molecular dynamics in a DMPC bilayer. The resulting ensemble revealed that the extracellular loops, which bind host receptors, occupy compact conformations, interact with each other weakly, and are dynamic on the nanosecond time scale. We predict that this conformational sampling is critical for enabling diverse Opa loop sequences to engage a common set of receptors.



INTRODUCTION

Neisseria gonorrheae (NG) and *Neisseria meningitides* are Gram-negative bacterial pathogens responsible for gonorrhea and meningococcal meningitis, respectively. For these bacteria, phagocytosis, cellular invasion, is induced by the binding of "opacity-associated" (Opa) proteins to host receptors. Opa proteins are eight-stranded β-barrel integral outer membrane proteins with four extracellular loops (Figure 1A). High sequence diversity is observed in regions of the extracellular loops of Opa variants (Table S1, Supporting Information), which is predominantly a result of recombination events between genes of the same isolate (70%) and import of genes from other isolates (16%).[1] The three variable regions within the extracellular loops, hypervariable 1 and 2 (HV1 and HV2) and the semivariable (SV) regions, engage host receptors to induce phagocytosis and determine the specific host receptors engaged.[2−4] The regions vary in length and do not comprise the entire extracellular loop: SV is 3−10 amino acids, HV1 is 24−31 amino acids, and HV2 is 45−51 amino acids. There are 26 SV, 96 HV1, and 127 HV2 different sequences in the 338 distinct opa alleles sequenced (http://www.neisseria.org). This sequence diversity likely plays a beneficial role in helping Neisseria to evade host immune responses;[5] however, it poses a challenge in that highly varied loop sequences must engage a common set of receptors to mediate cellular invasion. The determinants of Opa−receptor interactions are of the utmost importance for understanding Neisserial pathogenesis and the innate immune response.[5−7] Opa proteins also provide a means of foreign body cellular entry through specific human receptors that can be exploited synthetically for pharmaceutical and technological purposes.

Opa proteins are classified into two subgroups based on host receptor selectivity. Opa_{HS} bind to heparansulfate proteoglycans (HSPGs) directly and indirectly to integrin receptors via an HSPG-mediated interaction. The more abundant class, Opa_{CEA}, bind to the nonglycosylated face of the carcinoembryonic antigen-related cellular adhesion molecule (CEACAM) Ig N-domain.[8] Although all CEACAM receptors contain this domain, Opa proteins only bind to CEACAM1, 3, 5, and 6, and most selectively bind to only a subset.[2,5,9−11] The CEACAM N-domain residues that interact with Opa_{CEA} have been identified: Y34 and I91 are essential for all Opa_{CEA} interactions, and an additional seven nearby residues are implicated in binding depending on the particular Opa_{CEA}.[4] The specificity-determining residues on Opa_{CEA} are predominantly in the HV1 and 2 regions, and the HV sequences are concomitant: chimeric Opa proteins with an HV1 and HV2 region from two Opa proteins that bind the same receptor do not bind.[2]

Received: September 23, 2013
Revised: March 27, 2014
Published: May 9, 2014

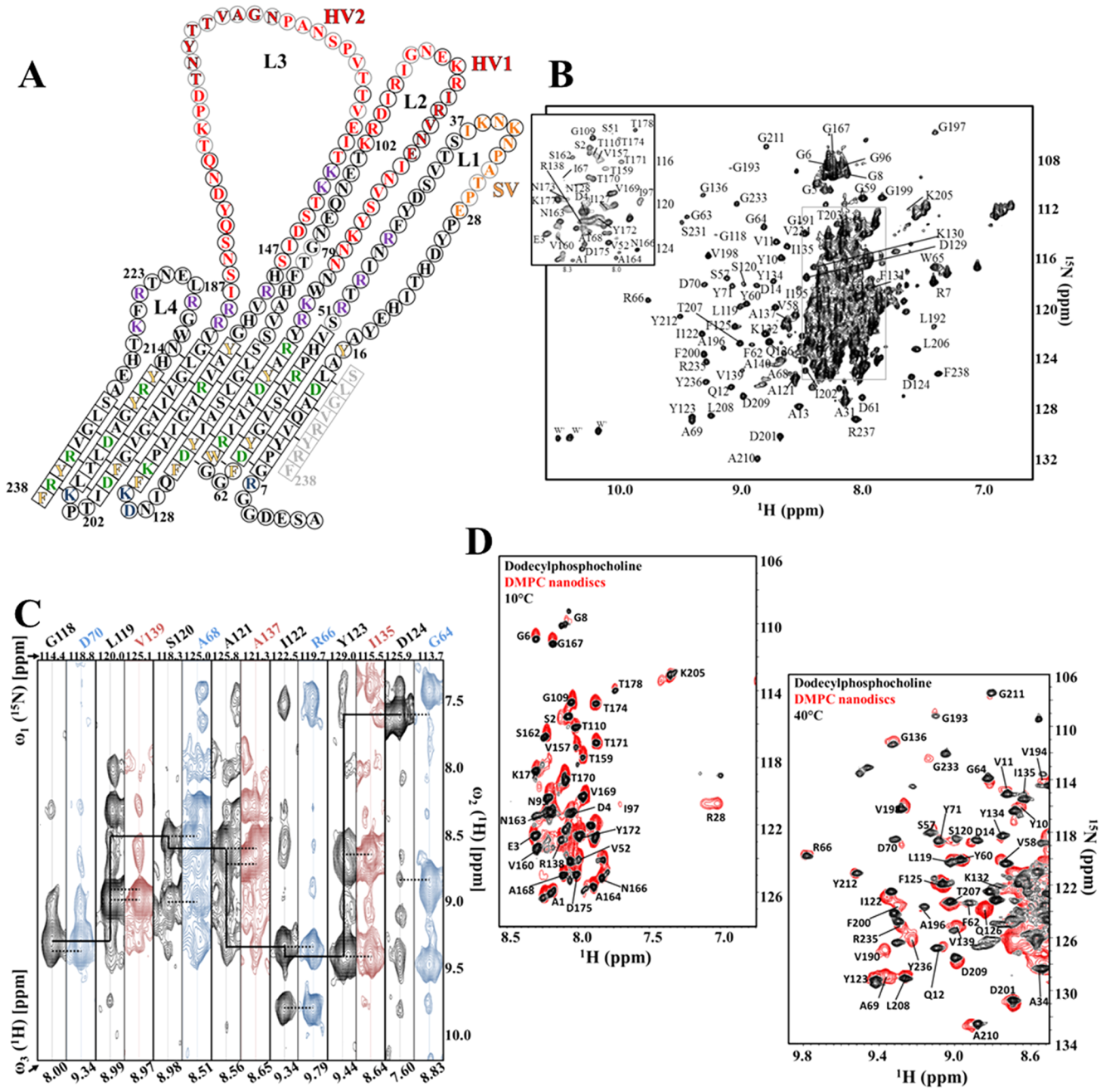


Figure 1. Opa_{60} topology and representative NMR spectra. (A) Opa_{60} is an eight-stranded β-barrel with long extracellular loops (~60% of the protein). Charged residues within the β-barrel region are colored green and in the periplasmic turns colored blue. Basic residues that may interact with LOS are colored purple. Aromatic residues located near the headgroup region of the bilayer are colored yellow. The semivariable region (SV) is colored orange, and the hypervariable (HV) regions 1 and 2 are colored red. Residues in the loops that have NMR assignments have gray circles and regions demonstrating transient helices in the MD simulations have black outlined font. Black lines indicate inter β-strand NOEs observed. (B) The ^{15}N, ^{1}H TROSY-HSQC is labeled with the NMR backbone assignment for Opa_{60} in DPC micelles. (C) Sample strips from the ^{15}N-NOESY spectrum indicating intra- (solid lines) and interstrand (dashed lines) NOEs observed. Residues in β-strand 4 (black), β-strand 3 (blue) and β-strand 5 (red) are labeled. (D) NMR ^{15}N, ^{1}H TROSY-HSQC spectra of Opa_{60} in nanodiscs containing DMPC lipids with peaks labeled with the NMR backbone assignment. The spectra shown were recorded at 800 MHz and at 40 °C (B, C, and D) or 10 °C to resolve only loop resonances (D). All NMR samples were deuterated with labile protons back exchanged and the protein concentrations were ~750 μM for dodecylphosphocholine samples and ~500 μM for nanodiscs preparations.

Toward understanding the molecular recognition required to gain entry into human cells, we report the structure of Opa_{60}, which binds CEACAM1, 3, 5, and 6.[2] Structure determination of membrane proteins is challenging and even more so for proteins that have large portions of both soluble and membrane-embedded regions. Thus, Opa proteins presented some methodological obstacles in both NMR resonance assignments[12] and structure calculation and refinement. To overcome these obstacles, we employed a hybrid method that used the restraints determined with solution nuclear magnetic resonance (NMR) spectroscopy in detergent micelles in conjunction with molecular dynamics (MD) simulations in a

dx.doi.org/10.1021/ja503093y | J. Am. Chem. Soc. 2014, 136, 9938−9946

lipid bilayer. This approach preserved the structural features that were well-determined spectroscopically but employed a more physical sampling method (molecular dynamics versus simulated annealing) and more detailed treatment of solvation and electrostatics to better define regions that either are flexible or remained underdetermined from the spectroscopic data alone. Opa_{60} is a canonical eight-stranded β-barrel with extensive ionic interactions inside the barrel. Three of the extracellular loops are longer than those found in any β-barrel structure previously determined. The HV regions within these loops are dynamic on the nanosecond time scale and are predominantly disordered. However, the loops are compact and interact with each other weakly such that long-lived specific intraloop interactions are not observed. The diverse and dynamic nature of the loop structural ensemble is likely required for highly variable Opa loop sequences to bind a common receptor and also for a single Opa protein to bind a variety of host receptors.

EXPERIMENTAL SECTION

Expression, Purification, and Refolding of Opa_{60}. Protocols for Opa_{60} expression, purification, and refolding were previously published.[13] The gene for Opa_{60} with N- and C-terminal fusion tags was subcloned into pET28B from the orginal pEX vector provided (Martine Bos, Ultrech University) and transformed into BL21(DE3) *E. coli*. Cells were grown in D_2O (99.8%) minimal media containing 4 g/L ^{13}C(99%)-glucose and 1 g/L ^{15}N(99%)-ammonium chloride (Cambridge Isotopes Laboratory) at 310 K until an OD_{600} of 0.8 expression was induced with 1 mM isopropyl-β-thio-D-galactoside for 8 h. Cells were lysed after resuspension in 50 mM Tris−HCl and 150 mM NaCl (lysis buffer). Cell debris from the lysate was removed via centrifugation at 12000*g* for 30 min. The pellet was resuspended in lysis buffer with the addition of 8 M urea overnight and centrifuged again at 12000*g* for 30 min. The soluble fraction was added to a Co^{2+}-immobilized metal affinity chromatography column and washed with 15 CV of 20 mM sodium phosphate, pH 7.8, 150 mM NaCl, 20 mM imidazole, 8 M urea followed by a 5 CV elution (20 mM sodium phosphate, pH 7.0, 150 mM NaCal, 680 mM imidazole). The eluted protein fraction was concentrated to 200 μM and rapidly diluted 20-fold with 20 mM Tris−HCl, pH 8.0, 500 mM NaCl, and 4.5 mM *n*-dodecylphosphocholine (DPC; Anatrace). After 5 days of room temperature incubation, the protein was fully folded as assessed with SDS−PAGE gel shift analysis. The sample was then concentrated and dialyzed against 3 × 4 L of 20 mM sodium phosphate, pH 6.2, and 150 mM NaCl for 1 h each. Final NMR samples were concentrated to 400−800 μM and contained 110−150 mM dodecylphosphocholine (DPC) as measured by comparing sample detergent intensities with standard concentrations.

Nanodisc Preparation and Opa Reconstitution. Opa_{60} was reconstituted into nanodisks according to established protocols[14−16] using plasmid for MSP1D1ΔH5 generously provided by Gerhard Wagner (Harvard University). MSP1D1ΔH5 was purified and assembled in 20 mM Tris/HCl pH 7.5, 100 mM NaCl, and 5 mM EDTA buffer with the appropriate amount of dry lipid/detergent to obtain a mixture of MSP1D1ΔH5/DMPC/sodium cholate with a molar ratio of 1:50:100. Opa_{60} refolded in DPC was added to the mixture, and the Opa_{60}/MSP1D1ΔH5 ratio was adjusted to 1:4. The mixture was incubated at 4 °C for 1 h, and detergent was removed with ∼0.5 g of washed Biobeads SM-2 (Biorad) per mL of assembly mixture. This suspension was gently agitated at 4 °C for 6−10 h. Biobeads were pelleted by centrifugation, and the decanted supernatant was concentrated and purified on a Superdex 200 gel filtration column equilibrated with 20 mM sodium phosphate, pH 6.5, 50 mM NaCl, and 5 mM EDTA. Fractions corresponding to the main peak were pooled and concentrated using an Amicon centrifugal filter unit of 30 kDa MWCO (Millipore). The NMR sample consisted of ∼0.5 mM $^{2}H,^{15}N$ Opa_{60} in MSP1D1ΔH5 nanodiscs with d_{54}-1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC; Avanti Polar Lipids), in gel filtration buffer supplemented with 10% (v/v) D_2O.

NMR Spectroscopy. NMR spectra were collected on Bruker AVANCE spectrometers operating at proton frequencies of 600 and 800 MHz equipped with Bruker 5 mm TXI cryoprobes and recorded at 313 K. Spectra were processed with Topspin and assigned using CARA (cara.nmr.ch). The assignment strategy for Opa_{60} is published[12] and mapped onto the $^{15}N,^{1}H$-TROSY-HSQC in Figure S1 (Supporting Information). Through these strategies, complete nitrogen, hydrogen, Cα, Cβ, and CO resonances were assigned for residues 1−14, 29−31, 51−71, 95−97, 109−110, 118−140, 157, 159−178, 190−212, 231, and 233−238 (Figure 1B) with only Cα, Cβ, and CO resonances for the seven assigned prolines and a lack of Cβ assignment for two additional resonances (Y71 and I97). Further, the entire side chain carbon and hydrogen assignments were obtained for residues 159−178 through TOCSY and COSY assignment using a corresponding synthetic peptide. Additionally, nine aromatic side chain protons (Y10, F62, W65, F125, F131, Y134, F200, Y236, and F238) were observed due to incomplete protein deuteration and dynamics of the side chain and assigned using the ^{15}N-NOESY spectrum. Resonance assignments were achieved for 92% of the β-barrel region (as defined by the MD refined structure) and 27% of the extracellular loops. The NspA structure was not used for the assignment or identification of restraints.

NMR Structure Calculations. The TALOS+ program was used to obtain backbone dihedral angle restraints. Assigned NOE peak heights were measured and binned into strong, medium, or weak interactions. Observed inter-β-strand NOEs are schematically indicated in Figure 1A. Thirty-one additional NOEs between backbone HN protons and aromatic side chain protons were also included. These were assigned upper limits of 3.5, 5.0, and 6.5 Å. In most cases, hydrogen bonding partners could be unambiguously assigned based on NOE patterns (a representative strip is shown in Figure 1C), and two distance restraints were used with upper limits of 2.5 and 3.5 Å for HN···O and N···O, respectively. Additionally, planar restraints (32 Å ± 4 Å as a square potential) were used to represent the lipid bilayer. The extracellular loops are very long, and without the bilayer restraint they sampled conformers that would be embedded into or transverse the bilayer (Figure S1, Supporting Information). The restraint distance was chosen on the basis of (1) the hydrophobic thickness of PorB a β-barrel membrane protein from NG for which there is a crystal structure[17] and (2) the residues that were observed to have NOEs with detergent headgroup choline protons (Figure S2, Supporting Information). The structure calculations were performed using NIH-XPLOR v2.31.[18,19] Starting at 3000 K, 5000 steps of high-temperature annealing was used to fold the initial extended structure. Twenty of the lowest overall violation energies of the 300 calculated structures with selected for further MD simulations.

MD Simulations. All simulations were performed using Gromacs 4.5[20] and the Charmm36 forcefield[21] for protein and lipid interactions. As detailed below, distance and dihedral restraints derived from the NMR data and used in XPLOR simulated annealing runs were enforced throughout the molecular dynamics simulations. Simulations were run under NPT conditions using the velocity-rescaling thermostat[22] at 300 K with a time-constant of 0.1 ps and semi-isotropic pressure coupling using the Parrinello−Rahman[23] barostat at 1 bar. All covalent bonds were constrained using LINCS,[24] and long-range electrostatics were computed every step using Particle Mesh Ewald (PME).[25]

A lipid bilayer of 512 DMPC molecules (bilayer thickness, 34 Å ± 1 Å;[26] hydrophobic thickness, 23−26 Å[26,27]) was constructed using the CHARMM-GUI membrane builder tool[28] and solvated with approximately 40000 TIP3P waters. Ions were added to obtain a system with 150 mM NaCl and no net charge. The DMPC bilayer was equilibrated prior to protein insertion with a resulting area per lipid headgroup of 0.60 nm^2, close to the experimentally determined value (0.606 ± 0.005 nm^2).[29] The system dimensions were approximately 12.5 nm (sides) and 11 nm (height).

Each of the 20 lowest energy structures from XPLOR simulated annealing was independently inserted in the equilibrated membrane

using the Gromacs tool g_membed,[30] removing approximately 15−20 lipids in the process. The five N-terminal protein residues were also removed. Each system was then energy minimized for 1000 steps using the steepest descents method. Twenty production runs, one per protein structure, were then carried out for 100 ns using a time step of 2 fs. Snapshots were saved every 500 ps, yielding 200 structures for each simulation. NOE-based distance restraints and TALOS+-derived dihedral restraints as used in the XPLOR simulated annealing calculations were imposed using spring potentials with force constants of 1000 kJ/mol/nm^2 and 1000 kJ/mol/rad^2, respectively.

To perform the Cα density analysis, all trajectories were aligned to a single consistent reference structure using a rigid-body fit where the objective function was calculated only on the β sheet and turn residues. Then the density was calculated on a 3D-grid with the MDAnalysis toolkit,[31] using a grid-spacing of 0.1 nm. The resulting density grid was visualized using PyMol (Schrödinger).

Clustering was performed using the Gromacs tool g_cluster, which was extended to include the *k*-means algorithm. Clustering was performed on all 4000 simulation snapshots to yield 50 clusters. The 20 most-populated, and therefore lowest free energy clusters, were selected, and the minimum energy structure from each was reported to form the hybrid refinement ensemble. Secondary structure analysis used the program DSSP[32] though the Gromacs tool do_dssp. The Cα RMSD within each cluster ranged from 1.52 to 4.22 Å and 1.87 to 5.33 Å for the entire protein and the extracellular loops, respectively (Table S2, Supporting Information). The RMSD for the extracellular loops between clusters is much greater with the pairwise RMSD for the minimum energy structures ranging from 5.60 to 29.3 Å (Table S3, Supporting Information). Only contacts greater than 1.5 kT estimated free energy difference between the MD and XPLOR structures were considered.

Additional analysis of contact between the HV1 and HV2 loops was performed by generating contact maps between these residues, where a contact was defined using a 6Å interatomic distance cutoff. Contact probabilities and lifetimes were computed using these contact maps, and highly contacting structures (defined as >50 simultaneous contacts) were further analyzed via agglomerative single-linkage clustering using Euclidean distance on the HV1-HV2 contact maps, yielding 10 clusters of HV1−HV2 contacts.

Calculation of Chemical Shifts. Chemical shifts were calculated on all snapshots in the MD simulations and for the XPLOR structures using the Sparta+ software.[33] The calculated shifts were then averaged for each atom, and the average values compared to the experimentally determined shifts (which were corrected for deuterium and TROSY induced shifts), where existing. In all cases, the errors reported are those used by Sparta+ (0.92 ppm for Cα and 0.49 ppm for HN, respectively).[33]

Backbone Dynamics. Relaxation rates were measured using two-dimensional ^{15}N−^{1}H TROSY-based experiments recorded at 600 MHz and 313 K. NMR data were processed and fit with NMRPipe.[34] Backbone dynamics were assessed by calculating the time autocorrelation function of the N−H bond vector $C_i(t) = \langle \mu_i(0), \mu_i(t) \rangle$, averaged across all trajectories, where μ_i is the N−H bond vector for the *i*th residue.

Data Deposition. The chemical shifts have been deposited in the BioMagResBank under the BMRB accession no. 19343. Atomic coordinates for the XPLOR and MD/XPLOR refined 20 conformers representing the structure of Opa$_{60}$ have been deposited in the PDB (PDB ID: 2MLH and 2MAF, respectively).

RESULTS AND DISCUSSION

NMR Structure. Opa$_{60}$ is an eight-stranded β-barrel with four extracellular loops (Figures 1A and 2A). The combination of a stable, membrane-inserted β-barrel domain and long unstructured loops complicated assignment and structure determination. Strategies for the assignment of the protein included trypsin cleavage, peptide synthesis, and assignment at various temperatures.[12] For the XPLOR structure determination, the β-barrel was calculated using the HN−HN amide proton and aromatic ring proton−HN NOEs, H-bond restraints, and backbone dihedrals calculated with TALOS (Table 1). The backbone RMSD of the barrel region for the 20 lowest energy structures is 0.96 Å (Table 1 and Figure 2A). The structure calculations were complicated by the long unstructured loops; in the initial calculation, extracellular loops adopted unreasonable conformations that spanned the membrane embedded region with excursions to the periplasmic side of the protein (Figure S1, Supporting Information). To address this problem, planar restraints were introduced at a distance of 32 Å ± 4 Å (see the Experimental Section for details).

Most of loops 1, 2, and 4 were not assigned (Figure 2B) because the resonances were not observed (although spectral

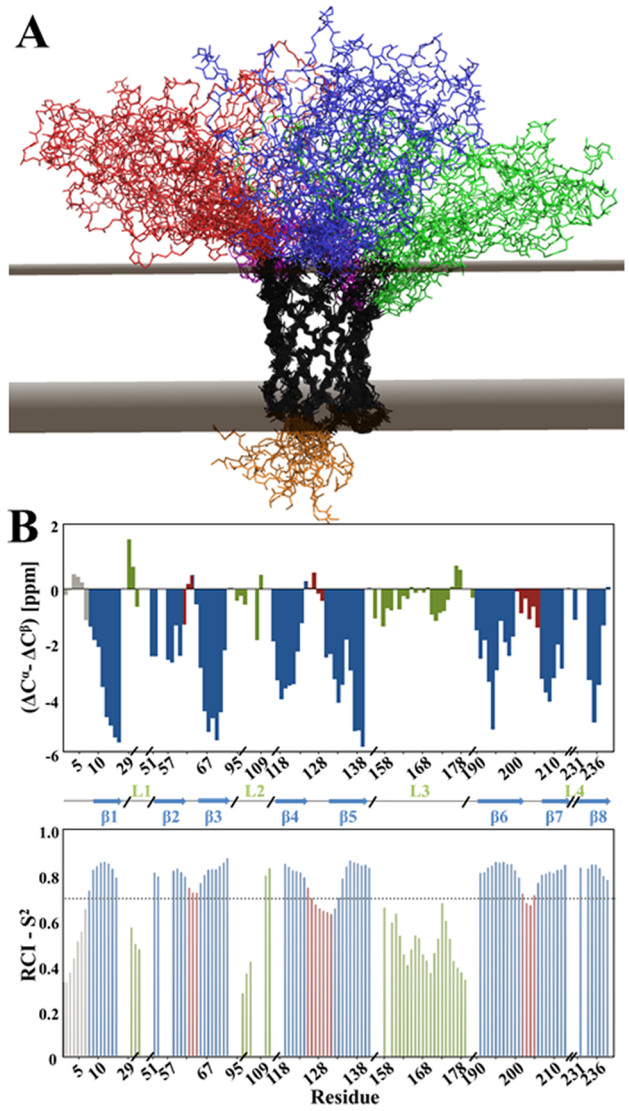


Figure 2. Opa$_{60}$ solution structure calculated with XPLOR. (A) β-Barrel and periplasmic turns are colored black; extracellular loop 1, green; loop 2, blue; loop 3, red; loop 4, magenta. The restraint planes are shown in brown. (B) Differences in carbon chemical shifts compared to random coil values are plotted; $(\Delta C^{\alpha} - \Delta C^{\beta}) = {}^{1}/_{3}(\Delta C^{\alpha}_{i-1} + \Delta C^{\alpha}_{i} + \Delta C^{\alpha}_{i+1} - \Delta C^{\beta}_{i-1} - \Delta C^{\beta}_{i} - \Delta C^{\beta}_{i+1})$. Order parameters calculated with the random coil index method.[35] Data for turns are colored red, for β-strands blue, for loops green, and for the N-terminus gray. The predicted secondary structure is shown as a schematic between the two graphs.

dx.doi.org/10.1021/ja503093y | J. Am. Chem. Soc. 2014, 136, 9938−9946

Table 1. NMR Opa_{60} Structure Statistics

	XPLOR	XPLOR/MD
PDB ID	2MLH	2MAF
HN-HN NOE	102	102
sequential	62	59
medium range	11	13
long range	29	30
HN-side chain NOE[a]	31	0
H-bond constraints	120	120
dihedral constraints	128	128
NMR constraint violations		
NOE (Å)	0.002	0.11
dihedral angle (deg)	0.106	1.04
NIH-XPLOR energy (kcal mol^{-1})	830.1 ± 76.7	
Ramachandran map analysis		
favored region (%)	78.3	90.4
allowed region (%)	11.9	7.2
outlier region (%)	9.8	2.2
ensemble RMSD		
mean global backbone RMSD (Å)		
β-sheet	0.84 ± 0.12	1.06 ± 0.23
β-sheet and turn	0.96 ± 0.11	1.19 ± 0.27
all	9.49 ± 1.73	7.97 ± 1.10
mean global heavy atom RMSD (Å)		
β-sheet	1.99 ± 0.20	2.56 ± 0.23
β-sheet	2.13 ± 0.18	2.65 ± 0.25
all	10.1 ± 1.6	9.48 ± 0.95

[a]Side chain restraints were not included in the MD/NMR hybrid refinement. XPLOR ensemble without side chain NOEs had a mean global backbone RMSD for the β-sheet residues of 1.04 ± 0.15 Å.

overlap contributed). The HV2 region was assigned using temperature and a synthetic peptide.[12] Lowering the temperature broadened β-barrel and some loop peaks beyond detection simplifying the spectra to only the most dynamic regions of the loops and facilitating 27 loop resonances to be assigned. To further assign the functionally important HV2 region, a synthetic 20 amino acid peptide was synthesized and had nearly complete spectral overlap with the full-length protein allowing 17 loop resonances to be assigned. The remaining loop resonances were significantly broadened and could not be assigned. There are two likely phenomena that contribute to the line broadening of the resonances that are not observed: (i) conformational exchange and (ii) structural heterogeneity. The former was speculated to contribute to the lack of assignments in OmpX[36] and OmpA[37] in dihexanoylphosphatidylcholine (DHPC) and DPC micelles, respectively. The missing resonances (approximately half of the extracellular loops) corresponded to residues that connect the ordered β-barrel and the flexible central region of the extracellular loops. Some of the missing loop resonances of OmpX were resolved when the micelle was replaced with nanodiscs containing DMPC and 1,2-dimyristoyl-*sn*-glycero-3-phosphoglycerol (DMPG).[14]

To better refine the structure, each of the 20 structures of the NMR ensemble was subjected to molecular dynamics simulations in a DMPC lipid bilayer. An interval of 100 ns of simulation was selected for the structure refinement because T1 and T2 relaxation data indicated the HV regions were dynamic on the nanosecond time scale. In addition, the chemical shifts and CD indicated the HV and extracellular loops, respectively, are random coil. Although the solution NMR structure was determined in DPC, the simulations were sought in a more biologically relevant membrane environment. There are several pieces of evidence that suggest Opa_{60} has the same structure in DMPC as in DPC. CD spectra of Opa_{60} in DPC and DMPC small unilamellar vesicles (liposomes) indicate the protein structure is approximately 50% random coil 50% β-strand (Figure S3, Supporting Information). In addition, Opa_{60} in nanodiscs with DMPC have β-barrel chemical shifts that are superimposable with the DPC spectra; however, a few barrel resonances corresponding to residues on strands 3, 6, and 8 are missing in the nanodisc spectrum (Figure 1D).

MD refined structure. To further refine the solution NMR structure, 100-ns MD simulations were performed on each of the 20 lowest energy NMR structures embedded in a DMPC lipid bilayer (bilayer thickness, 34 ± 1 Å;[26] hydrophobic thickness, 23−26 Å[26,27]). The resulting ensemble from the MD simulations has a backbone RMSD of 1.19 Å for the β-barrel region (Figure 3A and Table 1). The MD ensemble is

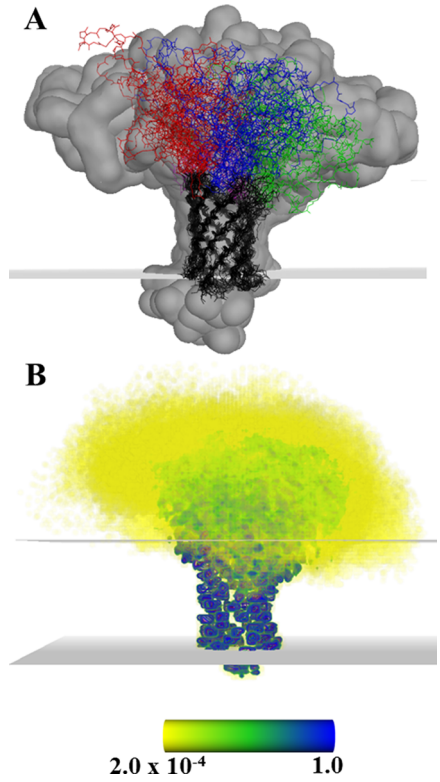


Figure 3. MD-refined Opa_{60} structure. (A) Structural ensemble colored as in Figure 1. The gray surface represents the volume occupied by the XPLOR loop ensemble. (B) Cα density over the 20 trajectories. Colored by density (probability of a Cα atom within an $Å^3$ grid cell) from yellow (2×10^{-4}) to blue. The gray planes show the average positions of the phosphorus atoms in the lipid head groups.

composed of the minimum energy structure from each of the 20 lowest free energy clusters and captures 2937 of the 4000 snapshots (Table 1). The Cα RMSD within each cluster ranged from 1.52−4.22 Å and 1.87−5.33 Å for the entire protein and the extracellular loops, respectively (Table S2, Supporting Information). The RMSD for the extracellular loops between clusters is much greater than within clusters, with the pairwise RMSD for the minimum-energy structures ranging from 5.60 to 29.3 Å (Table S3, Supporting Information). Although sampling

dx.doi.org/10.1021/ja503093y | J. Am. Chem. Soc. 2014, 136, 9938−9946

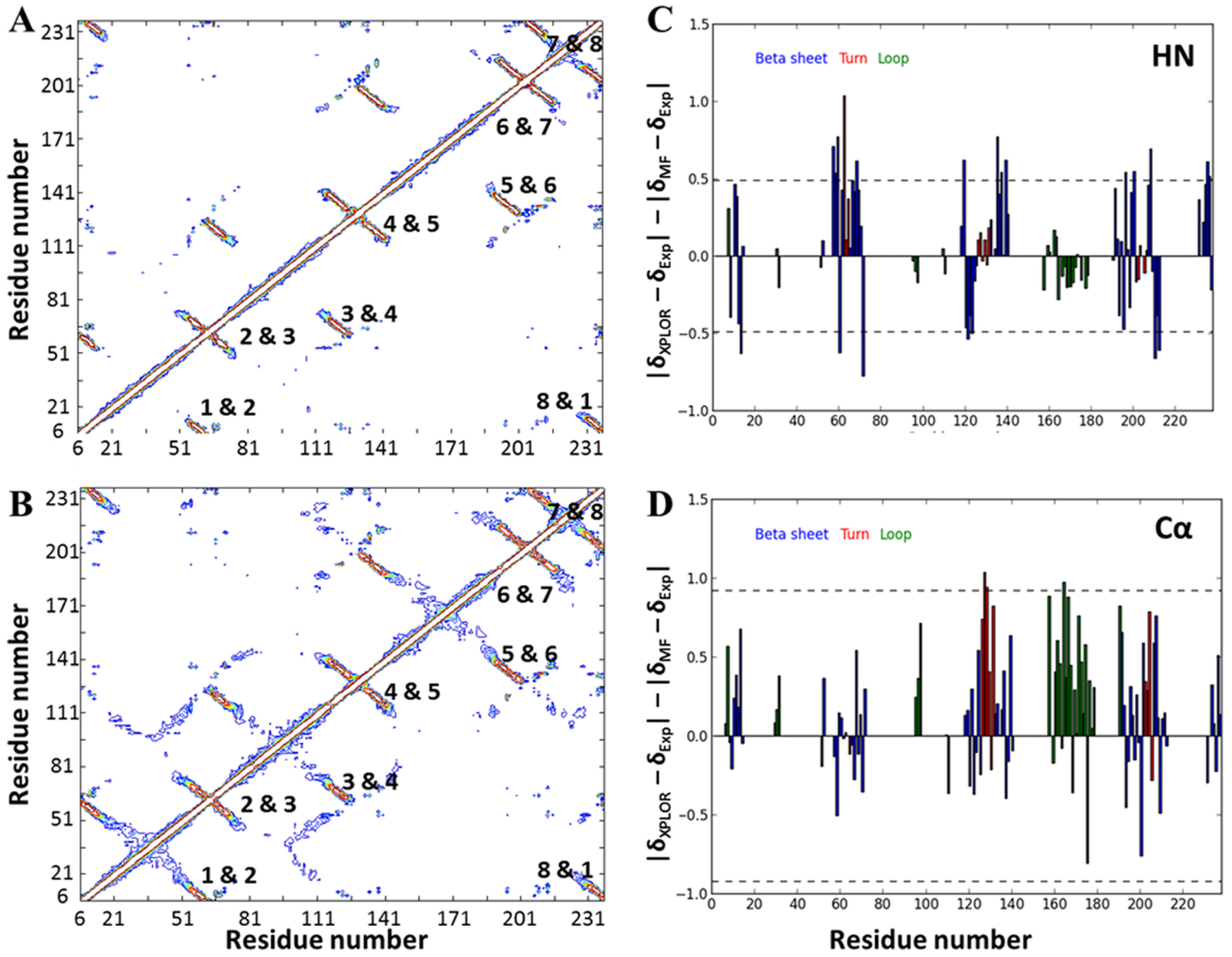


Figure 4. Comparison of XPLOR and MD refined Opa_{60} structures. Contact map for the ensemble calculated with XPLOR in detergent (A) and for the MD in lipid (B). The contact map is rendered as a contour plot of contact probability, with evenly spaced contours from 10% to 100% contact probability in each ensemble colored from dark blue to dark red. Contacts between β-strands are labeled. Contacts were defined as two atoms from respective residues approaching within 5 Å. Comparison of the amide proton (C) and Cα (D) chemical shifts calculated from the XPLOR ensemble and the MD ensemble. Positive values indicate the chemical shifts calculated from the MD ensemble agree better with the observed chemical shifts, and negative values indicate chemical shifts calculated from the XPLOR ensemble agree better with the observed chemical shifts. Dashed lines indicate Sparta+ prediction accuracy for each chemical shift.

is not sufficient to achieve full convergence, principal components analysis of the loop conformations across all simulations yields good overlap in the subspace of the two largest principal components (Figure S4, Supporting Information), indicating that at least in terms of the highest amplitude loop motions, individual simulations sampled overlapping rather than isolated regions of conformation space. Thus, the ensemble from the 20 lowest energy clusters represents the loop structural diversity observed in the trajectories.

The β-barrel of Opa_{60} is similar in sequence to NspA[38] (Figure S5, Supporting Information); however, the overall fold from the MD/NMR refinement is most similar to OmpA and OmpX (Table S4, Supporting Information). Similar to the 12 eight-stranded β-barrels deposited in the Protein Data Bank, the strands have a right-handed twist (Figure 3A) and an aromatic belt around the circumference of the barrel at the lipid headgroup regions (Figure S6, Supporting Information).[39,40] A significant ionic network exists inside the barrel (Figure S7, Supporting Information); however, there is no observable pore through the length of the barrel.

The ionic residues on the extracellular side of the barrel (Figure S7, Supporting Information) are excluded from solvent in many of the clusters, and those on the periplasmic surface (Figure S7, Supporting Information) are accessible to solvent in all the clusters. This ionic network may contribute to the significant stability observed for Opa_{60}; the barrel remains intact after cleavage with trypsin and boiling in SDS loading buffer.[12] There are additional ionic features in the Opa_{60} structure beyond the bilayer. Basic residues (Figure S8, Supporting Information) are clustered in the extracellular loops near the barrel domain. These residues may interact with the negatively charged lipooligosaccharide outer leaflet of the outer membrane as was observed for FhuA[41] and one of the LPS interactions identified with OprH.[42] However, specific interactions between the barrel and LPS were not detected by chemical shift perturbation when LPS (LOS, which is in the outer leaflet of the Neisseria outer membrane, is not commercially available) was titrated into the Opa_{60}–DPC micelle (data not shown).

Comparison of XPLOR and XPLOR/MD Structures. The MD-refined ensemble has several more inter-residue contacts than the XPLOR refined ensemble (Figure 4A,B). These new

dx.doi.org/10.1021/ja503093y | J. Am. Chem. Soc. 2014, 136, 9938–9946

contacts are primarily intraloop: 11 are across the periplasmic side of the β-barrel and 17 extend the strands on the extracellular side (only the 76 contacts that represented a greater than 1.5 kT estimated free energy difference between the MD and XPLOR structures were considered). These contacts position the loops more centrally above the barrel. Consequently, the most probable loop conformations in the MD refined structures are much more compact than the XPLOR refined structures (Figure 3A) with most of the loop density above the β-barrel (Figure 3B). This decrease in loop volume can be quantified via the protein radius of gyration, which decreases from the initial structure over the course of 19 of the 20 trajectories (Figure S9, Supporting Information).

The accuracy of the structural representation of the XPLOR and MD refined ensembles can be evaluated by comparing chemical shifts calculated from the ensemble structures via semiempirical shift prediction methods against the experimental chemical shifts (δ_{Exp}).[43] Parts C and D of Figure 4 compare the deviations between the experimentally measured values (δ_{Exp}) of Cα and HN Opa_{60} chemical shifts and the δ_{XPLOR} and δ_{MD} predictions. Most residues do not show differences between δ_{XPLOR} and δ_{MD} larger than the Sparta+ reported prediction accuracy. However, for carbon shifts, which primarily depend on backbone dihedral angles, the MD predictions agreed better with the experimental shifts indicating that the MD ensemble is an accurate representation of the observed Opa_{60} structure.

HV2 Dynamics. The β-barrel provides the scaffolding for the functional extracellular loops, which are disordered and sample a diverse ensemble of conformers. The NMR and MD dynamics data (Figure 5) indicate that HV2 and three extracellular loops (L1−L3), respectively, are dynamic on the nanosecond time scale. Because the Opa−detergent complex has a large overall correlation time, T1 values are highly sensitive to backbone nanosecond motions. Opa T1 values decrease significantly at the N-terminus, periplasmic turns, and extracellular loops 1 and 3 compared to the β-strands (Figure 5A), indicating these regions have high amplitude motions in the nanosecond time scale. Several of the β-strands have a general trend of increased dynamics toward the N- and C-terminal ends of the strands (although most have order parameters greater than 0.85), which was previously reported for other β-barrel membrane proteins investigated with NMR.[37,44] T2 changes are much more difficult to interpret since values increase with nanosecond motions and decrease with μs−ms motions. Nonetheless, the Opa_{60} T2 values (Figure 5A) are consistent with the interpretation of the T1 values. Thus, on the basis of the NMR data, the HV2 region of Opa has a high amplitude of motion on the nanosecond time scale. These NMR data are consistent with those observed with MD, which provides a more comprehensive understanding of the motions of the loops. Although sampling of loop conformations was not globally converged, the rank order of backbone dynamics showed good convergence. Assessed at 20 ns, the Spearman ρ was 0.975 between the mean autocorrelation function value and the fifth percentile of sampled trajectories, while the ρ between the mean and the 95th percentile was 0.988. A gradient is observed for the MD-derived backbone NH bond vector time autocorrelation function for each of the three longer extracellular loops (L1−L3; Figure 5B), with loop regions furthest from the barrel more dynamic than the regions closest to the barrel. Based on the MD simulations, the SV, HV1, and HV2 regions are moving within the nanosecond time regime and with a high amplitude of motion.

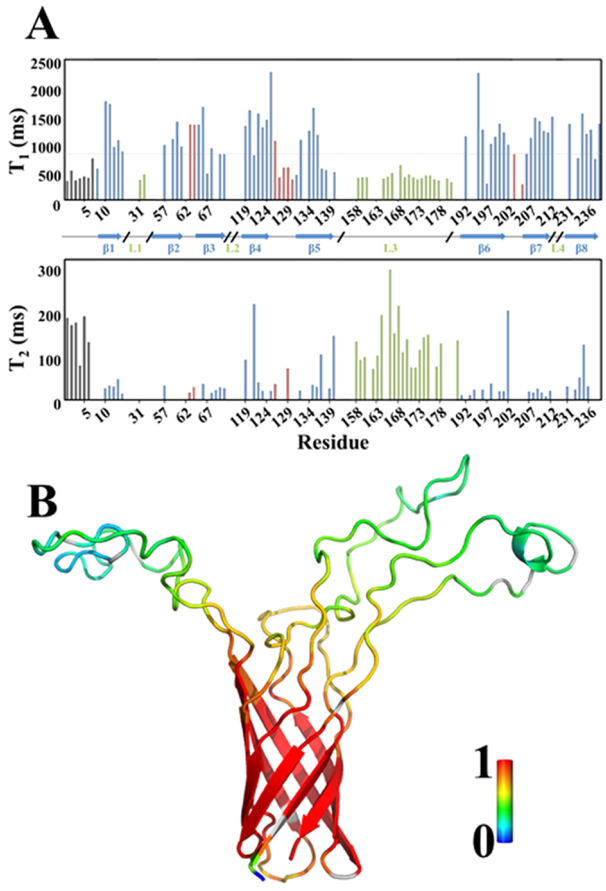


Figure 5. Dynamics of Opa_{60}. (A) ^{15}N T1 and ^{15}N T2 relaxation values for Opa_{60} plotted versus sequence and secondary structure. ^{15}N T1 values that are less than 80% (dotted line) of the value predicted (1 s) for a 20 ns overall correlation time have S^2 values less than 0.85 (τ_e = 10 ps). Data for turns are colored red; β-strands, blue; and loops, green. (B) Cartoon representation of Opa_{60} with residues colored according to the N−H orientational time autocorrelation function calculated from MD trajectories.

HV1 and HV2 Structure. As might be expected from the extensive nanosecond-time scale dynamics, the loops do not maintain long-lived structural features. However, they do form considerable intraloop contacts, which are captured more readily by the hybrid refinement strategy than simulated annealing alone (Figure 4A,B and 6A). There are several contacts within each loop that are observed in the MD refined

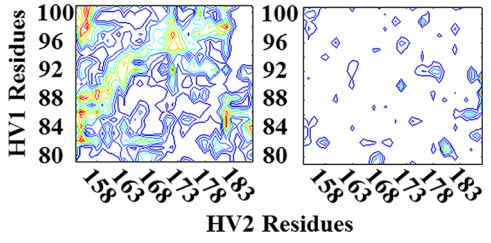


Figure 6. HV1−HV2 interactions. (A) HV1−HV2 contact map for all 4000 snapshots. Contacts were defined as two atoms from respective residues approaching within 6 Å. Each contour line represents a percentile increment in contact probability. (B) Average contact lifetimes are plotted for each contact, with each contour line representing a 5-ns lifetime increment. Comparison of these panels shows that HV1−HV2 contacts are relatively frequent but short-lived.

dx.doi.org/10.1021/ja503093y | J. Am. Chem. Soc. 2014, 136, 9938−9946

structures. Of the 76 contacts above 1.5 kT estimated free energy difference between the XPLOR and XPLOR/MD ensembles, 31 are within each of the three loops (L1−L3). Contacts between HV1 and HV2 are of most interest since they are both required to bind to CEACAM receptors. Common contacts between the two regions are observed (Figure 6A); however, these contacts have short lifetimes (Figure 6B). Throughout the simulations these contacts are frequent but short-lived. Based on this observation, the 4000 snapshots were reclustered based on the HV1 and HV2 regions and analyzed in terms of contacts and representative structures (Figure S10, Supporting Information). Recurrent contacts were observed between residues in the range 153−165 (HV1) and 86−95 (HV2) as well as 171−180 (HV1) and 94−98 (HV2). Not surprisingly, there are several hydrophobic residues that mediate these HV1−HV2 interactions (Figure S10, Supporting Information). The recurring contact patterns were observed in loop conformations that were globally quite different and across multiple independent MD simulations from different starting structures, suggesting robust formation of transient yet frequent interactions. These observations are broadly consistent with the primary NMR data in that long-lived interactions or persistent structure in HV1 and HV2 were not observed on the basis of chemical shift (Figure 2B) and the lack of nonsequential NOEs in assigned regions.

In addition to the contacts observed, the SV, HV1, and HV2 regions each sample helical conformers in a small fraction of the 4000 snapshots of the 20 trajectories (Figure S11, Supporting Information). Other secondary structures, such as PPII and 3_{10} helices, were less abundant in these regions (Figure S11, Supporting Information). The existence of these lowly populated structures is difficult to probe with traditional NMR methods; however, for populated secondary structure elucidation, carbon chemical shifts (Figure 2B) are typically used. For the data obtained, only a few residues in the SV and HV2 regions indicate α-helical structure (positive values), but overall the values indicate the dominant population is random coil which is consistent with the MD results. The lack of any long-lived discrete structure in the extracellular regions of Opa_{60} is compatible with the degree of sequence variability that still confers binding to host receptors (Table S1, Supporting Information). It would be surprising should such extreme variability result in a single stable structure.

Insights into Opa-Receptor Recognition. Despite the structural plasticity of the extracellular loops, Opa proteins must still bind a common set of receptors. Depending on the hypervariable sequences in the extracellular loops, Opa proteins bind selectively to the N-domain of CEACAM1, 3, 5, and/or 6 but do not bind the N-domains of CEACAM4, 7, and 8.[11] Using mutagenesis, residues Y34 and I91 of CEACAM N-domains (Figure S12, Supporting Information) were identified to be essential for the Opa-receptor interaction.[2,11] Several other residues (27, 28, 29, 32, 39, 44, and 89) dictate the different Opa−CEACAM selectivity reported (Figures S12 and S13, Supporting Information). The total exposed surface area of these identified residues is approximately 440 $Å^2$ [45] and is composed of both hydrophobic and polar moieties, which can easily be complemented by the hydrophobic and polar groups in the HV regions of Opa proteins (Table S1, Supporting Information). Beyond the enthaplic interactions, the dynamics and conformations of the extracellular loops are important to the molecular recognition event. The extracellular loops are intrinsically disordered yet are sampling a restricted volume such that there are interactions between the loops on the nanosecond time scale. MD simulations further suggest recurrent yet transient interaction patterns between specific regions of the loops. Thus, the loops adopt an intermediate state that is not folded but is also not lacking in interactions; the state of the loops may be best described as premolten globule or "fuzzy".[46] In order to bind CEACAM, both HV1 and HV2 are required; the HV regions are concomitant since chimeric Opa proteins with an HV1 and HV2 region from two Opa proteins that bind the same receptor do not bind receptor.[2] The "fuzzy" state may be a mechanism to retain disorder yet provide conformers in which HV1 and HV2 are in proximity and competent to interact with CEACAM.

The small CEACAM binding surface (Figure S13, Supporting Information) and the requirement of both Opa hypervariable regions suggest that a large folding event of the extracellular loops is unlikely upon binding suggesting that the binding mechanism is more likely conformational selection rather than induced fit. However, there is a plethora of commentary on the similarities and differences of these two binding mechanisms with the prevailing idea that binding reactions could have elements of both mechanisms.[47−49] In addition, sequences are selected for function not mechanism; therefore, the mechanism of different Opa proteins may vary. Since Opa−CEACAM interactions differ among variants and receptors, the dynamic nature of the loops maximizes the likelihood that a sequence will engage the receptor by increasing conformer sampling and the potential binding modes for receptor engagement.

■ CONCLUSION

We report the structure of Opa_{60}, a Neisserial outer membrane protein that induces host phagocytosis of the bacterium through specific receptor interactions. This eight-stranded β-barrel protein possesses three extracellular loops (greater than 34 residues) that are longer than any β-barrel structures yet reported and required a membrane restraint in the XPLOR structure calculation. To understand the structure and dynamics of the loops, we employed a hybrid XPLOR/MD refinement where NMR-derived restraints were used in 20 × 100 ns MD simulations to obtain a structure of Opa_{60} in a DMPC membrane. The hybrid-refined structure is consistent with the initial XPLOR structure but has an increase in loop structure and compactness. The loops are highly dynamic with backbone motions on the nanosecond time scale. Although there is little secondary structure evident in either the MD simulations or the primary NMR data, there are many short-lived contacts between the loops on the nanosecond time scale due to extension of the β-strands. We hypothesize that the observed dynamic ensemble is critical for maximizing the conformations of a highly variable region of Opa to engage receptors. That is, a high degree of plasticity is required to tolerate the diverse sequences in these regions and sample conformers competent to engage receptors. It remains to be shown, however, whether different Opa variants engage a single receptor (e.g., CEACAM1) via similar or different loop structures. The structure of several Opa variants in complex with an identical receptor will elucidate whether the binding modes are indeed convergent or divergent.

■ ASSOCIATED CONTENT

Supporting Information

Referenced supplemental figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

columbus@virginia.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Alison Criss for helpful discussions, Iza Bielnicka and Chris Reyes for technical assistance cloning, expressing, and purifying of Opa_{60}, Dr. Jeff Ellena for NMR technical support, and Ali Khan for assistance in the calculation of the NMR ensemble with bilayer restraints. Computational resources were provided by FutureGrid, NSF XSEDE, and Folding@Home donors. P.L. acknowledges a Marie-Curie postdoctoral fellowship (PIOF-GA-2010−275548). This research was supported by NIH RO1GM087828 (L.C.), NSF MCB 0845668 (L.C.), Cottrell Scholar Award, RCSA (L.C.), UVA nanoSTAR (L.C.), and NIH R01GM098304 (P.M.K.).

■ REFERENCES

(1) Bilek, N.; Ison, C. A.; Spratt, B. G. *J. Bacteriol.* **2009**, *191*, 1878.
(2) Bos, M. P.; Kao, D.; Hogan, D. M.; Grant, C. C.; Belland, R. J. *Infect. Immun.* **2002**, *70*, 1715.
(3) Grant, C. C.; Bos, M. P.; Belland, R. J. *Mol. Microbiol.* **1999**, *32*, 233.
(4) Virji, M.; Evans, D.; Hadfield, A.; Grunert, F.; Teixeira, A. M.; Watt, S. M. *Mol. Microbiol.* **1999**, *34*, 538.
(5) Sadarangani, M.; Pollard, A. J.; Gray-Owen, S. D. *FEMS Microbiol. Rev.* **2011**, *35*, 498.
(6) Criss, A. K.; Seifert, H. S. *Nat. Rev. Microbiol.* **2012**, *10*, 178.
(7) Normark, S.; Albiger, B.; Jonsson, A. B. *Nat. Immunol.* **2002**, *3*, 210.
(8) Billker, O.; Popp, A.; Gray-Owen, S. D.; Meyer, T. F. *Trends Microbiol.* **2000**, *8*, 258.
(9) Callaghan, M. J.; Jolley, K. A.; Maiden, M. C. *Infect. Immun.* **2006**, *74*, 5085.
(10) de Jonge, M. I.; Vidarsson, G.; van Dijken, H. H.; Hoogerhout, P.; van Alphen, L.; Dankert, J.; van der Ley, P. *Infect. Immun.* **2003**, *71*, 2331.
(11) Popp, A.; Dehio, C.; Grunert, F.; Meyer, T. F.; Gray-Owen, S. D. *Cell Microbiol.* **1999**, *1*, 169.
(12) Fox, D. A.; Columbus, L. *Protein Sci.* **2013**, *22*, 1133.
(13) Dewald, A. H.; Hodges, J. C.; Columbus, L. *Biophys. J.* **2011**, *100*, 2131.
(14) Hagn, F.; Etzkorn, M.; Raschle, T.; Wagner, G. *J. Am. Chem. Soc.* **2013**, *135*, 1919.
(15) Denisov, I. G.; Grinkova, Y. V.; Lazarides, A. A.; Sligar, S. G. *J. Am. Chem. Soc.* **2004**, *126*, 3477.
(16) Ritchie, T. K.; Grinkova, Y. V.; Bayburt, T. H.; Denisov, I. G.; Zolnerciks, J. K.; Atkins, W. M.; Sligar, S. G. *Methods Enzymol.* **2009**, *464*, 211.
(17) Zeth, K.; Kozjak-Pavlovic, V.; Faulstich, M.; Fraunholz, M.; Hurwitz, R.; Kepp, O.; Rudel, T. *Biochem. J.* **2013**, *449*, 631.
(18) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. *Progr. NMR Spectrosc.* **2006**, *48*, 47.
(19) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65.
(20) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845.
(21) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. *J. Phys. Chem. B* **2010**, *114*, 7830.
(22) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, 126.
(23) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182.
(24) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116.
(25) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(26) Lewis, B. A.; Engelman, D. M. *J. Mol. Biol.* **1983**, *166*, 211.
(27) Silvius, J. R. *Thermotropic Phase Transitions of Pure Lipids in Model Membranes and Their Modifications by Membrane Proteins*; John Wiley & Sons: New York, 1982.
(28) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. *Biophys. J.* **2009**, *97*, 50.
(29) Kucerka, N.; Liu, Y. F.; Chu, N. J.; Petrache, H. I.; Tristram-Nagle, S.; Nagle, J. F. *Biophys. J.* **2005**, *88*, 245a.
(30) Wolf, M. G.; Hoefling, M.; Aponte-Santamaria, C.; Grubmuller, H.; Groenhof, G. *J. Comput. Chem.* **2010**, *31*, 2169.
(31) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *J. Comput. Chem.* **2011**, *32*, 2319.
(32) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577.
(33) Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13.
(34) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.
(35) Berjanskii, M. V.; Wishart, D. S. *Nucleic Acids Res.* **2007**, *35*, W531.
(36) Fernandez, C.; Hilty, C.; Bonjour, S.; Adeishvili, K.; Pervushin, K.; Wuthrich, K. *FEBS Lett.* **2001**, *504*, 173.
(37) Arora, A.; Abildgaard, F.; Bushweller, J. H.; Tamm, L. K. *Nat. Struct. Biol.* **2001**, *8*, 334.
(38) Vandeputte-Rutten, L.; Bos, M.; Tommassen, J.; Gros, P. *J. Biol. Chem.* **2003**, *278*, 24825.
(39) Jackups, R.; Liang, J. *J. Mol. Biol.* **2005**, *354*, 979.
(40) Wimley, W. C. *Protein Sci.* **2002**, *11*, 301.
(41) Ferguson, A. D.; Hofmann, E.; Coulton, J. W.; Diederichs, K.; Welte, W. *Science* **1998**, *282*, 2215.
(42) Edrington, T. C.; Kintz, E.; Goldberg, J. B.; Tamm, L. K. *J. Biol. Chem.* **2011**, *286*, 39211.
(43) Robustelli, P.; Stafford, K. A.; Palmer, A. G., III. *J. Am. Chem. Soc.* **2012**, *134*, 6365.
(44) Hwang, P. M.; Choy, W. Y.; Lo, E. I.; Chen, L.; Forman-Kay, J. D.; Raetz, C. R.; Prive, G. G.; Bishop, R. E.; Kay, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13560.
(45) Fedarovich, A.; Tomberg, J.; Nicholas, R. A.; Davies, C. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 971.
(46) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2.
(47) Csermely, P.; Palotai, R.; Nussinov, R. *Trends Biochem. Sci.* **2010**, *35*, 539.
(48) Dyson, H. J. *Q. Rev. Biophys.* **2011**, *44*, 467.
(49) Espinoza-Fonseca, L. M. *Biochem. Biophys. Res. Commun.* **2009**, *382*, 479.