Crystal Structure of HAb18G/CD147

IMPLICATIONS FOR IMMUNOGLOBULIN SUPERFAMILY HOMOPHILIC ADHESION

Received for publication, April 8, 2008, and in revised form, April 22, 2008. Published, JBC Papers in Press, April 22, 2008, DOI 10.1074/jbc.M802694200

Xiao-Ling Yu1, Tianten Hu1, Jia-Mu Du1, Jian-Ping Ding1, Xiang-Min Yang1, Jian Zhang1, Bin Yang6, Xu Shen5, Zheng Zhang1, Wei-De Zhong1, Ning Wen*, Hualiang Jiang1, Ping Zhu1,2, and Zhi-Nan Chen1,3

From the 1Cell Engineering Research Center & Department of Cell Biology, State Key Laboratory of Cancer Biology, and the 3Department of Clinical Immunology, Xijing Hospital, the Fourth Military Medical University, 17 West Changle Road, Xi’an 710032, the 2Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, 555 Zu Chong Zhi Road, Shanghai 201203, the 4State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, 5Guangzhou First Municipal People’s Hospital, Guangzhou Medical College, Dongfeng Xi Road, Guangzhou 510182, and **Chinese People’s Liberation Army General Hospital, 28 Fuxing Road, Beijing 100853, China

CD147, a member of the immunoglobulin superfamily (IgSF), plays fundamental roles in intercellular interactions in numerous pathological and physiological processes. Importantly, our previous studies have demonstrated that HAb18G/CD147 is a novel hepatocellular carcinoma (HCC)-associated antigen, and HAb18G/CD147 stimulates adjacent fibroblasts and HCC cells to produce elevated levels of several matrix metalloproteinases, facilitating invasion and metastasis of HCC cells. In addition, HAb18G/CD147 has also been shown to be a novel universal cancer biomarker for diagnosis and prognostic assessment of a wide range of cancers. However, the structural basis underlying the multifunctional character of CD147 remains unresolved.

We report here the crystal structure of the extracellular portion of HAb18G/CD147 at 2.8 Å resolution. The structure comprises an N-terminal IgC2 domain and a C-terminal IgI domain, which are connected by a 5-residue flexible linker. This unique C2-I domain organization is distinct from those of other IgSF members. Four homophilic dimers exist in the crystal and adopt C2-C2 and C2-I dimerization rather than V-V dimerization commonly found in other IgSF members. This type of homophilic association thus presents a novel model for homophilic interaction between C2 domains of IgSF members. Moreover, the crystal structure of HAb18G/CD147 provides a good structural explanation for the established multifunction of CD147 mediated by homo/hetero-oligomerizations and should represent a general architecture of other CD147 family members.

CD147 (also known as EMMPRIN (extracellular matrix metalloproteinase inducer), Basigin, M6, neurothelin, and HAb18G) is a type I integral transmembrane glycoprotein that belongs to the immunoglobulin superfamily (IgSF). It plays pivotal roles in intercellular interactions involved in tumor metastasis and angiogenesis (1, 2), spermatogenesis and fertilization (3, 4), neural network formation and development (5, 6), and human immunodeficiency virus infection and rheumatoid arthritis (7, 8). CD147 is known to have acted as a central factor in the stimulation of matrix metalloproteinases (MMPs) required for tumor invasion and metastasis via cell-cell and cell-matrix interactions. Several studies have shown that the expression and functional activities of CD147 are associated with and regulated by other important molecules, including integrins (α3β1 and α6β1) (9), monocarboxylate transporters (MCT1 and MCT4) (10), and cyclophilins (CypA, CypB, and Cyp60) (7, 11, 12). It has also been reported that CD147 stimulates the production of vascular endothelial growth factor and hyaluronan, leading to angiogenesis and multidrug resistance, respectively (2, 13). It has widely been accepted that CD147 is a multifunctional transmembrane protein mediating molecular events that are crucial to many biological circumstances. However, the molecular mechanisms and structural basis whereby CD147 modulates numerous phenomena are still unclear.

HAb18G/CD147, a novel hepatocellular carcinoma (HCC)-associated antigen, was cloned in our laboratory by screening the HCC cDNA expression library using anti-HCC mAb HAb18 (14). Over the years, we have conducted comprehensive researches on this antigen and its specific antibody, mAb HAb18, both in vitro and in vivo. Our studies have found that HAb18G/CD147 stimulates adjacent fibroblasts and HCC cells to produce elevated levels of several MMPs, facilitating invasion and metastasis of the HCC cells (15). Our studies have also suggested that HAb18G/CD147 may be responsible for the increased MMP secretion, cell invasion, and CypA-mediated cell migrant into the joints, thus contributing to the cartilage and bone destruction of rheumatoid arthritis (8). HAb18G/CD147 may also function in facilitating the entry of host cells by
severe acute respiratory syndrome coronavirus, which confirms some previous reports that CD147 enhances human immunodeficiency virus-1 infection by interacting with virus-associated CypA (7). All our findings have indicated that HAb18G/CD147 is an important molecule in tumor progression, viral infection, and rheumatoid arthritis pathogenesis and thereby a potential drug target for therapeutic interventions. In light of this, we have developed specific mAb HAb18 against HAb18G/CD147 as a radioimmunotherapeutical agent, named Licartin (generic name, metuximab, iodine-131, for injection). Licartin has been proved to be safe and effective for targeted treatment of HCC in clinical trials (16) and was approved as a new drug for clinical therapy of primary HCC by the China State Food and Drug Administration (No. S20050039) in April, 2005.

Research has led to a better understanding of the biological functions of CD147, but the structural basis of CD147 underlying its multifunctional character remains to be elucidated, especially its homo-oligomerization and the heterophilic interactions with other biological molecules such as integrins, cyclophilins, and its mAbs. The lack of a three-dimensional structure of HAb18G/CD147 hampers the design of new antibodies and small molecule regulators against this antigen. Thus, defining the structure of HAb18G/CD147 is of great interest and pharmaceutical importance.

The x-ray crystal structure of the soluble extracellular portion of HAb18G/CD147, solved to 2.8 Å resolution, consists of an N-terminal IgC2 domain and a C-terminal Ig1 domain, which are connected by a 5-residue flexible linker. Although it has been predicted that CD147 belongs to the IgSF and is highly homologous to the Ig variable (V) domain and major histocompatibility complex class II β chain (17, 18), the C2-I domain organization of HAb18G/CD147 we revealed from the crystal structure is unique in comparison with all known structures of IgSF members, which usually adopt V or I set Ig fold at their N termini. This structure provides for the first time an atomic view of an IgSF member with a C2-I domain organization and is likely to represent a general architecture of other CD147 family members. Notably, four copies of HAb18G/CD147 molecules are present in the asymmetric unit with virtually identical individual domains but divergent interdomain orientations and buried surface areas, indicating that these molecules adopt different conformations due to the flexibility of the linker. Moreover, four different pairwise combinations of HAb18G/CD147 molecules were observed in the crystal. Taken together, the unique domain organization, overall flexibility, and diverse dimerization manners of HAb18G/CD147 provide structural clues for the multiple oligomerization-dependent functions of this molecule.

**EXPERIMENTAL PROCEDURES**

**Expression and Purification of Extracellular Portion of HAb18G/CD147**—A cDNA encoding amino acids 22–205 of HAb18G/CD147 was inserted into pET21a (+) (Novagen) with Ndel and XhoI, and the integrity was confirmed by automated sequencing. This construct was chemically transformed into the Escherichia coli strain and grown in LB, yielding secretion of soluble HAb18GEP. The bacterial pellet was resuspended and sonicated in 20 mM Tris-HCl, pH 8, and centrifuged at 18,000 × g for 30 min. The supernatant was applied directly to a HitTrap Q HP column, followed by a Mono Q ion exchange column (GE Healthcare). A Superdex 75 gel-filtration column (GE Healthcare) was used as the final purification step with 20 mM Tris-HCl, 150 mM NaCl, pH 8. The eluate was concentrated to 20 mg/ml for crystallization.

**Crystallographic data for HAb18G/CD147**—Tetragonal crystals were obtained with hanging drop vapor diffusion method in crystallization solution containing 0.5 M ammonium sulfate, 0.1 M trisodium citrate dihydrate, pH 5.6, 1.0 M lithium sulfate monohydrate at 4 °C. Crystals were cryoprotected in the mother liquor added with 30% of 6 M lithium acetate and flash-cooled in liquid nitrogen for data collection at 100 K. The native diffraction data were collected to 2.8 Å resolution at wavelength of 1.000 Å at beamline BL17A of the Photon Factory, Japan, using an ADSC Quantum 270 charge-coupled device detector. The single wavelength anomalous dispersion data were collected to 3.1 Å resolution at wavelength of 0.9792 Å at beamline NW12 of the Photo Factory using an ADSC Quantum 210 charge-coupled device detector. All data were indexed, integrated, and scaled using program HKL2000 (19). Crystals belong to space group P41212 containing four molecules in an asymmetric unit and have a solvent content of ~70%. Statistics for data collection are summarized in Table 1.

**Structure Determination and Refinement**—The structure of HAb18G/CD147 extracellular portion was solved using the single wavelength anomalous dispersion method as implemented in SOLVE (20). The single wavelength anomalous dispersion phases were improved by statistical density modification, including solvent flattening and histogram matching using the program RESOLVE (21), increasing the overall figure-of-merit from 0.30 to 0.62 at 3.1 Å resolution. RESOLVE automatically

| TABLE 1 | Data collection and refinement statistics |
|---------|-----------------------------------------|
| Native* | Selenomethionine*                       |
| Data collection | | |
| Space group | P4,2,2 | P4,2,2 |
| Cell dimensions | | |
| a, b, c (Å) | 90, 90, 90 | 90, 90, 90 |
| α, β, γ (°) | 126.5, 126.5, 169.9 | 125.2, 125.2, 169.7 |
| Wavelength | 1.0000 | 0.9792 |
| Resolution (Å) | 50.0–2.80 (2.90–2.80) | 50.30 (3.21–3.10) |
| Rmerge | 0.106 (0.586) | 0.114 (0.697) |
| I/σI | 18.4 (2.2) | 38.1 (5.8) |
| Completeness (%) | 99.4 (100) | 99.7 (100) |
| Redundancy | 7.3 (7.4) | 28.5 (29.6) |

* The number of crystals used for structure determination: 1 for native and 1 for selenomethionine.

* Values in parentheses are for the highest resolution shell.
Crystal Structure of HAb18G/CD147 and Implications

built a partial polyalanine model of 368 residues. A complete model of a single HAb18G/CD147 molecule was manually built in program O (22). This model was directly used as the search model to determine the initial phases of the 2.8 Å native data by molecular replacement method implemented in program CNS (23). Structure refinement was carried out with the standard protocols of CNS. At the final step of refinement, TLS parameters generated by using the TLSMD server and restrained refinement were used in program O (24). 95% data were used for structure refinement, and 5% data were used for monitoring the free R factor. The structure quality was evaluated by using the program PROCHECK (25). The Ramachandran plot indicated that 84.7%, 12.9%, 1.9%, and 0.5% of the backbone torsion angles were in most favored, additionally allowed, generously allowed, and disallowed regions, respectively. The statistics of the structure refinement and the quality of structure models are summarized in Table 1. Secondary structures were assigned by DSSP (26). The graphics were prepared by using PyMOL.5

Molecular Dynamics Simulations of HAb18G/CD147 Dimers—The crystal structures of dimers were taken as the starting points for molecular dynamics (MD) simulations. Each MD simulation was carried out using the AMBER suite of programs (version 8.0) with the parm99 force field (27). Each structure was prepared by using the xLeap module in AMBER, in which protons were added to the structure. All ionizable side chains were maintained in their standard protonation states at pH 7.0. The proteins were solvated in cubic box of water molecules, with a water thickness extending at least 10 Å apart from the protein surface. To avoid the instability that might occur during the MD simulations, the solvated system was subjected to minimization for 5000 cycles with protein restrained and followed by another 5000 cycles with the whole system relaxed. Then, the system was gradually heated from 0 K to 300 K during the first 60 ps by three intervals, followed by equilibrium for 80 ps under constant volume and temperature (constant particle number (N), volume (V), and temperature (T)) condition. Afterward, the system was switched to constant pressure and temperature (constant particle number (N), pressure (P), and temperature (T)) condition and equilibrated for 100 ps to adjust the system to a correct density. Finally, the production simulations were carried out in the absence of any restraint under constant particle number (N), pressure (P), and temperature (T) condition, and the data were collected. This protocol was applied to all of the simulation systems. All the MD simulations were performed using the parallel version of PMEMD in AMBER suit. The particle mesh Ewald method was employed to calculate the long range electrostatic interactions, whereas the lengths of the bonds involving hydrogen atoms were fixed with the SHAKE algorithm (28) of Miyamoto and Kollman (29). During the simulations, the integration time step of 2 fs was adopted, and structural snapshots were flushed every 500 steps (1 ps). The non-bonded cutoff was set to 10.0 Å, and the non-bonded pair list was updated every 25 steps. Each production simulation was coupled to a 300 K thermal bath at 1.0 atm pressure by applying the Berendsen algorithm (30). The temperature and pressure coupling constants were set to 2.0 and 1.0 ps, respectively.

On the basis of the equilibrated dynamic trajectory, the binding free energy of each dimer was calculated by using the MM-GBSA method encoded in the AMBER 8.0 program. Coordinates from the dynamic trajectory were used every 20 ps (5 snapshots out of 100 ps were processed), and the MM-GBSA calculation was performed on each of them using the AMBER 8.0 program. For each snapshot collected during the simulation, dimer binding free energy (ΔG_{binding}) was calculated using Equation 1 (31, 32),

$$\Delta G_{binding} = \Delta G_{dimer} - [\Delta G_{monomer1} + \Delta G_{monomer2}]$$  \hspace{1cm} (Eq. 1)

where ΔG_{dimer}, ΔG_{monomer1}, and ΔG_{monomer2} are the free energies of the dimer and two monomers. Each free energy term in Equation 1 in AMBER 8.0 was calculated with the absolute free energy of the species (two monomers and their dimer) in gas phase (E_{gas}), the solvation free energy (ΔG_{solvation}), and the entropy term (TΔS) using Equation 2,

$$E_{gas} = E_{int} + E_{vdw} + E_{electrostatic}$$  \hspace{1cm} (Eq. 3)

$$E_{int} = E_{bond} + E_{angle} + E_{torsion}$$  \hspace{1cm} (Eq. 4)

The solvation free energy, ΔG_{solvation}, is approximated as the sum of the polar contribution (G_{GB}) and nonpolar contribution (G_{nonpolar}) using a continuum representation of the solvent in Equation 5.

$$\Delta G_{solvation} = G_{GB} + G_{nonpolar}$$  \hspace{1cm} (Eq. 5)

The polar contribution (G_{GB}) to the solvation energy was calculated using GB model in sander module. The nonpolar contributions (G_{nonpolar}) were estimated using a simple equation: G_{nonpolar} = γ × SASA + b kcal/mol (33). SASA is the solvent-accessible surface area that was estimated using the MSMS algorithm with a probe radius of 1.4 Å (34). The surface tension proportionality constant γ and the free energy of non-polar salvation for a point solute b were set to 0.00542 kcal/(molÅ²) and 0.092 kcal/mol, respectively.

The entropy calculation is extremely time-consuming for large systems. In addition, the main aim of calculating the binding free energy is to address the influence of mutations to the binding affinity. In principle, site-directed mutation may not result in dramatically conformational changes for the protein. Accordingly, entropy contributes little to the relative binding free energy changes for the binding strength of one dimer in comparison with other dimers. Therefore, in this study, ΔG_{binding} (without the term of −TΔS) was estimated to address the relative binding strength of the protein.
RESULTS

Overall Structure of the Extracellular Portion of HAb18G/CD147—HAb18G/CD147 crystallizes in space group P4₁,2,2 with four monomers in the asymmetric unit (designated A, B, C, and D) (Fig. 1). All residues are traced from the electron density except the first residue at the N terminus and 3–4 residues at the C terminus. The HAb18G/CD147 structure comprises two immunoglobulin-like domains connected by a five-residue linker. This architecture agrees with the previous prediction that CD147 is a member of the IgSF. Structurally, the N-terminal Ig-like domain (D1, residues 22–101) strongly resembles the C2 set fold, whereas the C-terminal one (D2, residues 107–205) shows strong homology to Ig variable domain (V set). Each domain adopts similar conformations among the four molecules in the crystal. The r.m.s.d. values from superposing the Cα atoms are 0.6–0.7 Å and 0.6–0.9 Å for D1 and D2, respectively. However, the relative orientation between the two domains is highly divergent, because the elbow angle (D1 mass center–hinge residue Pro-104–D2 mass center) varies from 116.4° to 163.0° (Fig. 2B).

Four different pairwise combinations of HAb18G/CD147 are detected in the crystal, joined mainly via edge-by-edge packing and association of β-sheets (Fig. 1). Three of these dimers involve molecules within the asymmetric unit (AC, BC, and AD dimers), and one is formed by crystallographic 2-fold symmetry (DD' dimer) (Fig. 1). All three potential glycosylation sites deduced from HAb18G/CD147 sequence are distant from the dimer interfaces (Fig. 2A), thereby eliminating the possibility of their interference with the dimerization, which is in accordance with the finding that the glycosylation of CD147 is irrelevant with its oligomerization (35). Chain A has the strongest electron density and thus will be used hereafter for analysis unless explicitly noted.

Structures of the Domains—The N-terminal domain (D1) is a typical C2 set immunoglobulin domain consisting of a β-barrel formed by sheets EBA and GFCC' and a conserved disulfide bond between strands B and F (Fig. 2A). Remarkably, none of the N-terminal Ig domains of the cell surface molecules with known structures have been discovered to adopt C2 Ig fold. Structural superpositions suggest that the overall size (~38 Å in longest axis) and strand topology of D1 are similar to those of the C2 domains from other cell surface glycoprotein antigens, including CD2 (~40 Å, PDB entries 1HNF and 1HNG) (36, 37), CD4 (~40 Å, PDB entries 1CDH and 1CID) (38, 39), and CD58 (~40 Å, PDB entry 1CCZ) (40). However, owing to the reduced size of its GFCC' sheet, D1 is smaller than the C2 domains from cell adhesion molecules, such as VCAM (~48 Å, PDB entry 1VCA) (41), ICAM-1 (~49 Å, PDB entry 1IAM) (42) and ICAM-2 (~49 Å, PDB entry 1ZXQ) (43). Noticeably, the EBA sheet is more negatively charged than the opposite GFCC' sheet. The loops flanking the EBA sheet are rich of conserved acidic residues (Fig. 3). Among them, the EF loop is most negatively charged, in which a conserved asparagine-rich DXDD motif forms a short 3₁₀ helix locating at the tail of the D1 and presumably pointing toward the cell membrane.

The C-terminal domain (D2) falls into the category of the I set. The β-sheets DEBA and A'GFCC' pack against each other, and a disulfide bond between Cys-126 and Cys-185 connects strands B and F together (Fig. 2A). Because the strands A and C' are rather short, D2 could also be...
regarded as a shortened V-like Ig domain, as Vaughn and Bjorkman have suggested (44). D2 (39 Å in longest axis) is of similar size as the heavy-chain variable domain of Fab NEW (35 Å in longest axis, PDB entry 7FAB (Fig. 4A)). In Fab, the CC and FG loops, also known as β-bulges, which introduce twists to strands C and G, are believed to be crucial to antibody V domain dimerization (45). In comparison, the F-G β-bulge of HAb18G/CD147 is substantially shorter and results in a less twisted AGFCC sheet. Meanwhile, the C-C bulge (Thr-143 to Asp-147) adopts a similar conformation as the C-C loop of Fab, but is rich of conserved acidic residues (DXX(D/E)) (Fig. 4A). In addition, the C-C bulge is one of the most flexible regions in D2, because the r.m.s.d. value from superposing this region is as high as 1.57 Å within the asymmetric unit.

Structural superposition of the two domains of HAb18G/CD147 revealed their striking resemblance. Although these two domains belong to distinct Ig domain types and share a sequence identity as low as 21%, their three-dimensional structures are very similar, with an r.m.s.d. value of 1.68 Å for 68 aligned Cα atoms (Fig. 4B). The intersheet spaces, strand topologies, disulfide bond positions, and even the conformations of some loops (especially the B-C loop whose sequence is conserved between D1 and D2) are quite constant. These similarities imply that the two Ig domains might have evolved from one common ancestral Ig domain. Their difference, however, in addition to the lack of strands A and D in D1, mainly occurs at the (A)GFCC sheet side, where strands F and G are slightly shorter in D1, and the characteristic C-C β-bulge of D2 is absent in D1 (Fig. 4B).

The potential N-linked glycosylation site, Asn-44, lies at the end of strand B, i.e. the outermost position of the EBA sheet. The other two glycosylation sites, Asn-152 and Asn-186, locate at the middle of C and strand F, respectively, with their side chains protruding oppositely from AGFCC and DEBA planes (Fig. 2A). This evenly distributed glycosylation sites around D2 might enable the molecule to stand upright on the
cell surface and fully expose the functioning surface of D1 for binding physiological ligands, as observed in ICAM-2 (46).

**Linker and Interdomain Geometry**—The domain arrangement of the four molecules in the asymmetric unit is quite different from each other: chains A and B are most bended, chain C is most extended, and chain D is in an intermediate state. This diversity of interdomain geometry is mainly caused by the 5-residue flexible linker containing a GPP motif (residues 103–105) that is conserved among all mammal CD147s (Fig. 3) and even exists in the far-related KIRR3 protein (Kin of IRRE-like protein, 28% sequence identity with that of HAb18G/CD147). Because proline and glycine are typically associated with polypeptide turns, this motif is believed to introduce the obvious bend observed between D1 and D2. More detailed inspection reveals that in chains A and B, the GPP motifs form hydrogen bonds (H-bonds) with the Ig domains, whereas these H-bonds disappear in chains C and D. This observation implies that the overall conformational variation among the four monomers may be associated with the H-bonds making and breaking between GPP motifs and the two domains. Supposing the last strand of D2 (strand G that connects to the transmembrane region) is perpendicular to the cell surface, the flexible linker might enable the EBA and GFCC’ sheets of D1 to sweep within a large area and form putative functional surfaces at the top of the molecule for ligand binding (Fig. 2, A and B) as seen in CD2 (36).

**Molecular Association in the Crystal**—Four different pairwise combinations of CD147 are observed in the crystal, representing possible dimer forms, and designated by their chain names as BC, AC, AD, and DD’ dimers. Interestingly, these dimers further adhere to each other by sharing some conserved β-strands located at either edge of the β-barrels, leading to homo-oligomer formation in the crystal lattice (Figs. 1 and 5).

**BC Dimer**—Many Ig domain-containing cell surface molecules, such as mouse and human JAM1 (47, 48), CD2 (36), CD4 by Val-61, Leu-62, Lys-63, Leu-67, and Phe-74 contact each other through hydrophobic interactions. In addition, the dimer is further stabilized by two H-bonds between C’ strands (Lys-63/B – Lys-63/C) and two salt bridges between E strands (Glu-73/B – Glu-73/C and Lys-75/B – Glu-73/C) (Fig. 5A, right). Consequently, the two GFCC’ sheets assemble into a large eight-strand β-sheet (GFCC’-C’CFG). The formation of the BC dimer is distinct from either the classic V-domain dimerization or CD2 adhesion interaction (36, 50), but resembles the D1-D1 dimerization of ICAM-1 (42), though its D1 domain belongs to I set.

From the sequence alignment of CD147 with neuroplastins (NPTN) and embigans from various species (Fig. 3), we could find two clusters of conserved residues involved in the dimer interface. The first cluster includes Val-61 and Leu-62 of strand C’. Val-61 is conserved among all mammal CD147s, and Leu-62 is completely conserved even in neuroplastins and embigans. This implies that the hydrophobic interaction between the two C’ strands and flanking loops is essential for the dimerization of HAb18G/CD147. The second cluster comprises Glu-73, Phe-74, and Lys-75. In CD147s from human, bovine, rabbit, Chinese hamster, as well as neuroplastins and embigans, the residues at Glu-73 position are usually occupied by acidic residues, and those corresponding to Lys-75 are always basic residues, suggesting that salt bridges analogous to those observed in HAb18G/CD147 could probably be formed between E strands in the homo-dimer. The position of Phe-74 is always occupied by an aromatic residue (phenylalanine or tyrosine) except the histidine in rabbit CD147. In BC dimer, Phe-74 is hydrophobically interacting with the conserved residue Leu-67 from the opposite monomer. Moreover, the potential N-glycosylation sites (Asn-44) from the two D1 domains locate far away from the dimer interface, further supporting the rationality of this dimer form, because glycosylation was found to have no effect on CD147 oligomerization (35).
all the dimerizations were calculated by using MD simulations. For the BC dimer, the calculated binding free energy was 

\[-20.84 \pm 4.74 \text{ kcal/mol}\]

(supplemental Table S1), which also indicates the existence of this dimerization.

**AC Dimer**—Chains A and C dimerize through their D1 domains with a buried surface area of 1004.6 Å² (Fig. 5B, left). In contrast to the BC dimer, the two EBA sheets of these two monomers constitute a large twisted six-strand \(\beta\)-sheet. The dimer interface is stabilized by H-bonds and hydrophobic interactions contributed by residues conserved among all mammal CD147s (Gly-24/A, Thr-25/A, Val-26/A, Thr-28/A, Thr-28/C, Val-30/A, and Val-30/C) (Fig. 5B, right). The assumption that the AC dimer might represent another form of transcellular homophilic interaction between two HAb18G/CD147 molecules is supported by the following observations: First, the two D2 dimers extend in opposite directions, and the overall length of the dimer is \(-103 \AA\), comparable with the assumed cell-cell distance. Second, the simulated binding free energy for this dimer \((-16.18 \pm 4.62 \text{ kcal/mol})\) is close to that of the BC dimer \((-20.84 \pm 4.74 \text{ kcal/mol})\) (supplemental Table S1). Third, the N-glycosylation site (Asn-44), although locating near the dimer interface, are pointing their side chains away from the interface into the bulk solvent and thus are not likely to interfere with the dimer formation.

**AD Dimer**—The D1 domain of chain D dimerizes with the D2 domain of chain A, forming an unusual C2-I interaction interface (Fig. 5C, left). Strand A of D1 from chain D adjoins antiparallel to strand G of D2 from chain A, resulting in a twisted EBA-GFCC'-B-sheet. The contact surface area of this dimer is \(-855.5 \AA²\). There are six H-bonds between the two domains (Fig. 5C, right). The binding free energy predicted for this dimerization is \(-29.69 \pm 2.22 \text{ kcal/mol}\) (supplemental Table S1), which is the largest among all observed dimers. In this case, the three potential N-glycosylation sites (Asn-44, Asn-152, and Asn-186) are inevitably positioned outside the dimer interface. This dimerization model implies that, in addition to homooligomerization (e.g. BC and AC dimers), the D1 domain of HAb18G/CD147 might also be involved in heterophilic interaction with other IgSF cell surface molecules that commonly use I/V-like Ig domains for ligand binding.
DISCUSSION

The unique C2-I domain arrangement distinguishes HAb18G/CD147 from all IgSF family members with known structures. The x-ray crystal structures of many IgSF adhesion molecules, such as CD4, CD2, VACM-1, ICAM-2, and ICAM-1, have been determined and characterized as an I-C2 domain organization (42). These cell surface molecules tend to utilize C2 set Ig domains as C-terminal building blocks to project the functional N-terminal domain (I/V-like Ig domain) away from the cell membrane (42). But little has been reported about the function of the C2 domain other than its supportive role. To our knowledge, the present study provides, for the first time, the three-dimensional structure of an IgSF molecule with an IgC2-IgI domain arrangement. Because of the special structural architecture of HAb18G/CD147, the role of the C2 domain, consequently, has shifted from a “supportive” to a “functional” one, which results in the C2-C2 (BC, AC, and DD dimers) and C2-I (AD dimer) dimerization in the crystal rather than V domain dimerization. Because classic V domain dimerization (50) and CD2 homophilic adhesion (36) have been regarded as typical paradigms for IgSF-IgSF adhesive interactions, the C2 domain-mediated dimerization presented in HAb18G/CD147 crystal structure might provide a novel example adding to the existing IgSF adhesion models. These C2-C2 and C2-I oligomerizations observed in the crystal also offer good structural clues for the established important functions of CD147 mediated by homo/heterophilic adhesive interactions, such as counter-receptor binding, MMP induction, and integrin association.

Four HAb18G/CD147 monomers in the asymmetric unit associate with each other forming three possible dimers, namely BC, AC, and AD. It is obvious that the four molecules adhere together by sharing monomer C between BC and AC dimers and by sharing monomer A between AC and AD dimers. The extensive intermolecular interactions observed in the crystal provide structural evidence for the nature of CD147 as an adhesion molecule of IgSF. Moreover, we presume that an adhesive “network” could probably be formed on the cell surfaces via tandemly repeated dimeric units of CD147, thereby facilitating its oligomerization-dependent multifunctions. Among these dimers, both BC and AC dimers are found to pack “head-by-head” via their N-terminal domains with end-to-end distances comparable to the assumed distance between two opposing cell surfaces. Thus, it is very likely that these dimers represent a trans-cellular homophilic interaction between two HAb18G/CD147 molecules on adjacent cell surfaces. Besides, numerous biological studies have emphasized the critical role of oligomerization in the multiple functions of CD147. Direct evidence has been provided for the presence of homo-oligomers of CD147 in chicken and mouse tissues by chemically cross-linking the membrane fraction (35, 51). Yoshida et al. have found that the oligomer of CD147 occurs in a cis-depend-ent manner on the plasma membrane and the N-terminal Ig-like domain is essential and sufficient for oligomerization, probably via hydrophobic interactions (35). Sun et al. have shown that CD147 may be a counter-receptor for itself to induce MMP production, and its trans-homophilic adhesion is mediated by the N-terminal domain in either heterotypic (tumor cell-fibroblast) or homotypic (tumor cell-tumor cell) circumstances. It is also suggested that the cis-multimerization of CD147 on the cell surface could enhance the overall avidity of its trans-interactions (52). Altogether, these biological evidences strongly suggest that the oligomerization-dependent biological functions of CD147 may require both cis and trans homophilic interactions on the surface of plasma membrane. In good agreement with this point, both cis and trans dimers were found in the crystal lattice of HAb18G/CD147, and the dimer...
Crystal Structure of HAb18G/CD147 and Implications

models we described could provide structural insights into the oligomerization of this molecule and supply a basis for mapping functionally critical regions as potential drug targets. Similar structural analysis about cis and trans oligomerizations of IgSF molecules could be found in the crystallographic study of human and mouse junctional adhesion molecules, in which the researchers deduced that the cis-dimers enhance and probably occur prior to any adhesive trans interaction (47, 48).

In addition to the homophilic interactions, CD147 has been reported to associate with integrins (49) that occur prior to any adhesive research (10). The human and mouse junctional adhesion molecules, in which the dimer BC. Thus the antibody might interfere with the mutagenesis study, partially overlaps with the interface of the epitope, predicted by molecular docking and confirmed by structural clues for drug discovery (supplemental Figs. S1 and S2). Interestingly, as indicated in supplemental Fig. S1C, the epitope, predicted by molecular docking and confirmed by mutagenesis study, partially overlaps with the interface of dimer BC. Thus the antibody might interfere with the oligomerization-dependent activities of HAb18G/CD147 that are critical for tumor metastasis, especially secretion of MMPs. This finding suggests that blocking the oligomerization of the molecule could be a reasonable strategy for developing more potent antibodies and small molecule antagonists against HAb18G/CD147 for treatment of hepatocellular carcinoma.

Acknowledgments—We thank the staff at the Photon Factory, Japan, for help with the data collection, and other members of our group for technical assistance and helpful discussion.

REFERENCES

1. Zucker, S., Hymowitz, M., Rollo, E. E., Mann, R., Conner, C. E., Cao, J., Foda, H. D., Tompkins, D. C., and Toole, B. P. (2001) Am. J. Pathol. 158, 1921–1928
2. Tang, Y., Nakada, M. T., Kesavan, P., McCabe, F., Millar, H., Rafferty, P., Bugelski, P., and Yan, L. (2005) Cancer Res. 65, 3193–3199
3. Igakura, T., Kadomatsu, K., Kaname, T., Muramatsu, H., Fan, Q. W., Miyauchi, T., Toyama, Y., Kuno, N., Yasuda, S., Takahashi, M., Senda, T., Taguchi, O., Yamamura, K., Arimura, K., and Muramatsu, T. (1998) Dev. Biol. 194, 152–165
4. Saxena, D. K., Oh-Oka, T., Kadomatsu, K., Muramatsu, H., and Toshimori, K. (2002) Reproduction 123, 435–444
5. Schlosshauer, B. (1991) Development 113, 129–140
6. Fadool, J. M., and Linser, P. J. (1993) Dev. Dyn. 196, 252–262
7. Pushkarsky, T., Zybarth, G., Dubrovsky, L., Yurchenko, V., Tang, H., Guo, H., Toole, B., Sherry, B., and Bukrinsky, M. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 6360–6365
8. Zhu, P., Ding, J., Zhou, J., Dong, W. J., Fan, C. M., and Chen, Z. N. (2005) Arthritis Res. Ther. 7, R1023–R1033
9. Berditchevski, F., Chang, S., Bodorova, J., and Hemler, M. E. (1997) J. Biol. Chem. 272, 29174–29180
10. Kirk, P., Wilson, M. C., Heddle, C., Brown, M. H., Barclay, A. N., and Halestrap, A. P. (1991) Biochem. J. 278, 786–788
11. Misra, S., Ghatak, S., Zoltan-Jones, A., and Toole, B. P. (2003) J. Biol. Chem. 278, 25285–25288
12. Chen, Z. N., Yang, Z., Mi, L., Ji, G., and Guo, X. (1999) J. Cell. Mol. Immunol. 15, 34
13. Xu, J., Xu, H. Y., Zhang, Q., Song, F., Jiang, L. J., Yang, X. M., Mi, L., Wen, N., Tian, R., Wang, L., Yao, H., Feng, Q., Zhang, Y., Xing, J. L., Zhu, P., and Chen, Z. N. (2007) Mol. Cancer Res. 5, 605–614
14. Xu, J., Shen, Z. Y., Chen, X. G., Zhang, Q., Bian, H. J., Zhu, P., Xu, H. Y., Song, F., Yang, X. M., Mi, L., Zhao, Q. C., Tian, R., Feng, Q., Zhang, S. H., Li, Y., Jiang, J. L., Li, L., Yu, X. L., Zhang, Z., and Chen, Z. N. (2007) Hepatology 45, 269–276
15. Miyauchi, T., Masuzawa, Y., and Muramatsu, T. (1991) Biochem. Biophys. Res. Commun. 188, 770–774
16. Fossum, S., Mallett, S., and Barclay, A. N. (1991) Eur. J. Immunol. 21, 671–679
17. Otwinowski, Z. M., and Minor, W. (1997) Methods Enzymol. 276, 307–326
18. Winn, M., Isupov, M., and Murshudov, G. N. (2001) Acta Crystallogr. D
Crystal Structure of HAb18G/CD147 and Implications

Ikemizu, S., Sparks, L. M., van der Merwe, P. A., Harlos, K., Stuart, D. I., Jones, E. Y., and Davis, S. J. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 4289–4294

Jones, E. Y., Harlos, K., Bottomley, M. J., Robinson, R. C., Driscoll, P. C., Edwards, R. M., Clements, J. M., Dudgeon, T. J., and Stuart, D. I. (1995) Nature 373, 539–544

Bella, J., Kolatkar, P. R., Marlor, C. W., Greve, J. M., and Rossmann, M. G. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 4140–4145

Casasnovas, J. M., Springer, T. A., Liu, J. H., Harrison, S. C., and Wang, J. H. (1997) Nature 387, 312–315

Vaughn, D. E., and Bjorkman, P. J. (1996) Neuron 16, 261–273

Chothia, C., and Lesk, A. M. (1985) J. Mol. Biol. 182, 151–158

Wang, J., and Springer, T. A. (1998) Immunol. Rev. 163, 197–215

Kostrewa, D., Brockhaus, M., D’Arcy, A., Dale, G. E., Nebbock, P., Schmid, G., Mueller, F., Bazzoni, G., Dejana, E., Bartfai, T., Winkler, F. K., and Hennig, M. (2001) EMBO J. 20, 4391–4398

Prota, A. E., Campbell, J. A., Schelling, P., Forrest, J. C., Watson, M. J., Peters, T. R., Aurrand-Lions, M., Imhof, B. A., Derody, T. S., and Stehle, T. (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 5366–5371

Ikemizu, S., Gilbert, R. J., Fennelly, J. A., Collins, A. V., Harlos, K., Jones, E. Y., Stuart, D. I., and Davis, S. J. (2000) Immunity 12, 51–60

Springer, T. A. (1991) Nature 353, 704–705

Fadool, J. M., and Linser, P. I. (1996) Biochem. Biophys. Res. Commun. 229, 280–286

Sun, J., and Hemler, M. E. (2001) Cancer Res. 61, 2276–2281

Tang, W., and Hemler, M. E. (2004) J. Biol. Chem. 279, 11112–11118

Becker, J. W., Erickson, H. P., Hoffman, S., Cunningham, B. A., and Edelman, G. M. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 1088–1092