JVI Journals.ASM.org

Structural Analysis of the Evolutionary Origins of Influenza Virus Hemagglutinin and Other Viral Lectins

Lang Chen, Fang Li

Department of Pharmacology, University of Minnesota Medical School, Minneapolis, Minnesota, USA

Influenza virus and other viruses use host cell surface sugars as receptors. Here we show that the sugar-binding domains in influenza virus hemagglutinin and other viral lectins share the same structural fold as human galectins (host lectins). Unlike the easily accessible sugar-binding sites in human galectins, the sugar-binding sites in viral lectins are hidden in cavities. We propose that these viral lectins originated from host lectins but have evolved to use hidden sugar-binding sites to evade host immune attacks.

The influenza virus imposes a major health threat on humans. Like many other sugar-binding viral glycoproteins (viral lectins), the influenza virus hemagglutinin (HA) uses sugar moieties on host cell membranes (e.g., glycoproteins, glycolipids, and glycoaminoglycans) as its receptor component for host cell entry. Hence, it is a major determinant of the host range, tropism, and antigenicity of the influenza virus. The crystal structure of influenza virus HA was determined over 3 decades ago (1), but its evolutionary origin remains unresolved. Tracking down the evolutionary origins of influenza virus HA and other viral lectins addresses a basic evolutionary question about viruses.

To date, crystal structures have been determined for the following viral lectins: rotavirus VP4 (2), adenovirus galectin domain (GD) (3), coronavirus spike protein N-terminal domain (NTD) (4, 5), coronavirus hemagglutinin-esterase (HE) (6), and torovirus hemagglutinin-esterase (7), in addition to influenza virus HA. Among them, rotavirus VP4, adenovirus GD, and coronavirus spike NTD have been shown to share the same structural folds with human galectins (host lectins), whereas influenza virus HA, coronavirus HE, and torovirus HE appear to have no structural homology with human galectins (based on protein structure database search server DALI [8]). Here, by investigating their structural topologies (connectivity of secondary structural elements) (9), we show that all of these viral lectins share the same structural folds with human galectins, suggesting that they all originated from a host galectin.

Human galectins contain a β-sandwich core structure consisting of one 6-stranded and one 5-stranded β-sheet (Fig. 1A and D). All of the viral lectins also contain a β-sandwich core. Among them, the β-sandwich cores of rotavirus VP4, adenovirus GD, and coronavirus spike NTD have the same structural topologies as human galectins except that coronavirus spike NTD has two more β-strands in one of the β-sheet layers (Fig. 1B and E). Compared with coronavirus spike NTD, the β-sandwich cores of influenza virus HA, coronavirus HE, and torovirus HE lack two β-strands in each of the β-sheet layers (Fig. 1C and F). Despite these structural differences, virtually all of the β-strands in these viral lectins are connected in the same order from the N terminus to the C terminus as human galectins (Fig. 1A to C). These results suggest that these viral and host lectins likely have gone through either convergent or divergent evolutionary paths to acquire related structural topologies, which will be discussed further in this article.

Despite their related structural topologies, these viral lectins and human galectins use different mechanisms to bind sugars. The sugar-binding site in human galectins is located on the top of the β-sandwich core (Fig. 2A). It is wide open and easily accessible to incoming sugars. The sugar-binding sites in viral lectins are hidden in cavities. Sugars are bound between the β-sheet layers in rotavirus VP4 (Fig. 2B), between the dimer interface in adenovirus GD (Fig. 2C), and in a pocket on one side of the β-sandwich core in influenza virus HA, coronavirus HE, and torovirus HE (Fig. 2D). The structure of a sugar-bound coronavirus spike NTD is not available, but mutagenesis studies have identified the sugar-binding site on the top of the β-sandwich core, which overlaps with the sugar-binding site in human galectins (5) (Fig. 2E). However, different from human galectins, the sugar-binding site in coronavirus spike NTD is underneath a ceiling structure that is the extension of loops connecting the β-strands in the core structure. It appears that viral lectins, but not host galectins, have undergone substantial evolution to come up with a variety of strategies to hide their sugar-binding sites.

Why do viral lectins need to hide their sugar-binding sites? The "canyon hypothesis" suggests that human rhinoviruses, which use host proteins as receptors, hide their receptor-binding sites in deep canyons to evade host immune surveillance (10). The cavities containing the sugar-binding sites in viral lectins are shallower and thereby more accessible to host antibodies than the canyons in rhinoviruses. However, compared with the sugar-binding site of human galectins, these cavities in viral lectins still significantly reduce the binding affinity and/or limit the accessibility of host antibodies to the sugar-binding sites (11). Here we hypothesize that as host proteins, human galectins are not recognized by the host immune system, whereas as foreign proteins, viral lectins need to hide their sugar-binding sites from host immune attacks.

How did these viral lectins originate and evolve? Because of the different sugar-binding mechanisms used by viral lectins and human galectins, sugar binding cannot be the selective pressure for a functionally convergent evolution of these viral lectins. In other

Received 18 December 2012 **Accepted** 19 January 2013
Published ahead of print 30 January 2013
Address correspondence to Fang Li, lifang@umn.edu.

doi:10.1128/JVI.03476-12

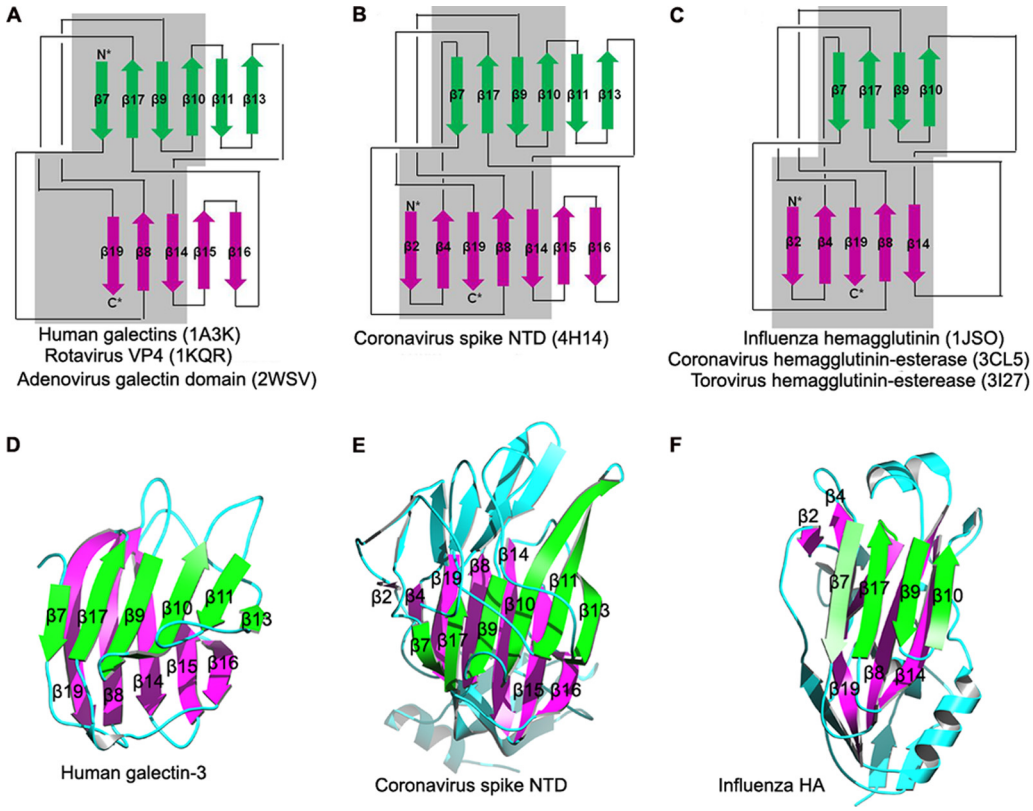

FIG 1 Structural comparisons among viral lectins and human galectins. (A) Structural topologies of human galectins, rotavirus VP4, and adenovirus galectin domain. (B) Structural topology of coronavirus spike NTD. (C) Structural topologies of influenza virus HA, coronavirus HE, and torovirus HE. The β-strands are named according to the coronavirus spike NTD structure (4). A common subcore structure is in gray. (D to F) Crystal structures of human galectin-3 (D), coronavirus spike NTD (E), and influenza virus HA (F).

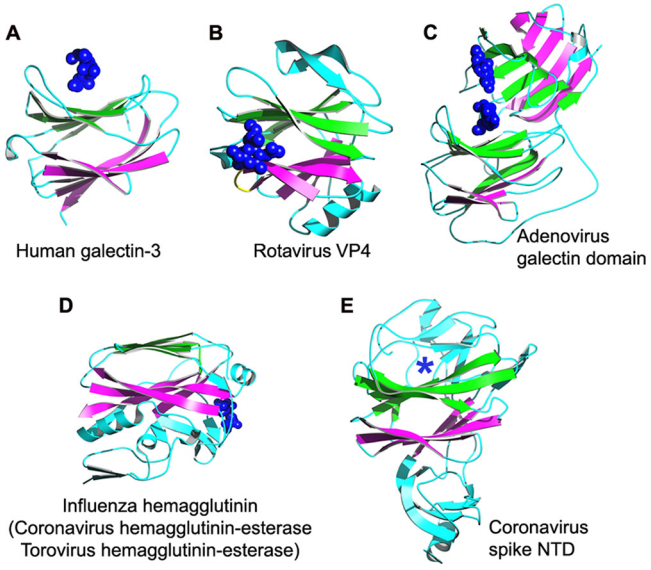

FIG 2 Sugar-binding sites in viral lectins and human galectins. The two β-sheets are in green and magenta. (A to E) Crystal structures of sugar-bound human galectin-3, rotavirus VP4, adenovirus galectin domain, influenza virus HA, and coronavirus spike NTD. Sugars are in blue. The asterisk indicates the sugar-binding site in coronavirus spike NTD that was identified by mutagenesis studies.

words, viral lectins would not be able to undergo convergent evolution to acquire related tertiary structures without a common sugar-binding mechanism as the evolutionary driving force. Instead, it is more likely that viral lectins originated from an ancient host lectin and have since diverged in their sugar-binding mechanisms. There may be more than one mechanism for the transfer of the lectin gene from hosts to viruses. One possibility is that one ancestral virus acquired the host lectin gene and all contemporary viral lectins evolved from this ancestral viral lectin. Alternatively, it is possible that different viruses independently acquired their lectin gene from hosts. Whatever the gene transfer mechanism is, once acquired by viruses, viral lectins have further evolved to adapt to viral host ranges and tropisms and to evade host immune surveillance.

To sum up, this study reveals two important evolutionary strategies used by influenza virus and other viruses, stealing a host lectin as their own cell entry machinery and evolving a variety of hidden sugar-binding sites to evade host immune attacks. The method of structural topology analysis used in this study may be useful to solve other evolutionary conundrums related to viral protein structures.

ACKNOWLEDGMENTS

This work was supported by NIH grant R01AI089728 (to F.L.).

Computer resources were provided by the Basic Sciences Computing Laboratory of the University of Minnesota Supercomputing Institute.

REFERENCES

1. **Wilson IA, Skehel JJ, Wiley DC.** 1981. Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature **289**: 366–373.
2. **Dormitzer PR, Sun ZYJ, Wagner G, Harrison SC.** 2002. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. EMBO J. **21**:885–897.
3. **Guardado-Calvo P, Munoz EM, Llamas-Saiz AL, Fox GC, Kahn R, Curiel DT, Glasgow JN, van Raaij MJ.** 2010. Crystallographic structure of porcine adenovirus type 4 fiber head and galectin domains. J. Virol. **84**:10558–10568.
4. **Peng GQ, Sun DW, Rajashankar KR, Qian ZH, Holmes KV, Li F.** 2011. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. Proc. Natl. Acad. Sci. U. S. A. **108**:10696–10701.
5. **Peng GQ, Xu LQ, Lin YL, Chen L, Pasquarella JR, Holmes KV, Li F.** 2012. Crystal structure of bovine coronavirus spike protein lectin domain. J. Biol. Chem. **287**:41931–41938.
6. **Zeng QH, Langereis MA, van Vliet ALW, Huizinga EG, de Groot RJ.** 2008. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. Proc. Natl. Acad. Sci. U. S. A. **105**: 9065–9069.
7. **Langereis MA, Zeng QH, Gerwig GJ, Frey B, von Itzstein M, Kamerling JP, de Groot RJ, Huizinga EG.** 2009. Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases. Proc. Natl. Acad. Sci. U. S. A. **106**:15897–15902.
8. **Holm L, Sander C.** 1998. Touring protein fold space with Dali/FSSP. Nucleic Acids Res. **26**:316–319.
9. **Li F.** 2012. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. J. Virol. **86**:2856–2858.
10. **Rossmann MG.** 1989. The canyon hypothesis—hiding the host-cell receptor attachment site on a viral surface from immune surveillance. J. Biol. Chem. **264**:14587–14590.
11. **Skehel JJ, Wiley DC.** 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. **69**:531–569.