Structure
Previews

CellPress

Interplay between intrinsic flexibility and sugar coating in Contactin-2 homodimerization

Mehmet Sen[1,*]
[1]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA
*Correspondence: msen2@cougarnet.uh.edu
https://doi.org/10.1016/j.str.2023.12.005

In this issue of *Structure*, Chataigner et al. reveal that Contactin-2's homotypic interaction, a glycosylation-dependent process, generates a broad conformational landscape. This structural plasticity, driven by conformational equilibria and sugar coating, facilitates adaptation to diverse ligands and environmental conditions, highlighting its dynamic role in neuronal function.

The Contactin subfamily, a group of IgSF (immunoglobulin superfamily) cell adhesion molecules, has six members. They contain an extracellular domain composed of six Ig-like domains at the N terminus and four fibronectin type-III repeats at the C terminus, linked to the plasma membrane through a glycosylphosphatidylinositol (GPI) anchor. Contactin-2, also called Axonin-1/TAG-1, is the functionally and structurally well-characterized member of the Contactin subfamily. Contactin-2, often via homotypic interaction, proficiently mediates cell-cell contacts in *trans* between molecules resident on apposed cell surfaces. Additionally, Contactin-2 is known to form sixteen heterotypic interactions both in *cis* and *trans* configurations. The physiological functions of Contactin-2 are reflected in guiding axon growth, promoting synapse formation, and contributing to the establishment of neural circuits. The precise roles of each Contactin-2 can vary in different contexts and developmental stages.

Structural views of Contactin-2, like many IgSF adhesion molecules, have been extensively explored through diverse biochemical and biophysical methods. The N-terminal Ig1-Ig4 domains of Contactin-2 adopt a U- or horseshoe-shaped structure akin to other IgSFs (e.g., Contactins 3, 4, 5, and 6, Dscam, and DCC).[1] These Ig1-Ig4 domains in Contactin-2 play a pivotal role in driving both heterotypic interactions (e.g., binding to NgCAM) and homotypic interactions, influencing neural cell surface adhesion and cell aggregation.[2] Crystallography studies revealed unique configurations for the C terminals of Contactin-3, Contactin-1, and -6, extended and bent, respectively. Additionally, FnIII domains were identified as facilitators of homotypic interactions. Two major models have been proposed to define the structural arrangement of Contactin-2 on cell surfaces: the zipper-like model, where *trans* homotypic interaction involves a loop in the Ig2-domain, and the four-molecules mode model, where Ig1-4 and FnIII domains participate in *trans* and *cis* interactions, respectively.

In this issue of *Structure*, Chataigner et al. employ an integrative approach, incorporating various techniques (in-solution X-ray scattering, multiangle light scattering, negative stain EM, X-ray crystallography, mass spectrometry, and mutational analysis) to examine and probe the conformational plasticity of Contactin-2.[3] The crystal structure of its Ig1-6 domains captures three molecules, all adopting the classic horseshoe conformation within their Ig1-4 domains. Previous size exclusion chromatography demonstrated the monodispersity of the Contactin-2 ectodomain, which adopts a trimeric or extended linear conformation.[4] A comparison of the two interacting dimers among the trimeric molecules in the crystal lattice reveals flexions between Ig domains, facilitating the formation of multiple conformations in the crystal lattice. Notably, L1, a related adhesion molecule of Contactin-2, was observed in the extended state.[5] The observed flexibility reveals diverse homotypic Ig1-Ig2 and Ig3-Ig6 dimers formed by striking conservation and massive surface areas within both interfaces. While Ig1-Ig2 dimerization strongly supports the zipper-like homotypic interaction, the Ig3-6 dimer also offers plausible insights into Contactin-2 oligomer formation.

Understanding the molecular basis of homotypic and heterotypic interactions within the Contactin family, along with the signaling events associated with each binding mode, is essential for unraveling the complex molecular mechanisms governing many physiological events involving Contactins. Various modes of interaction have indeed been observed in IgSFs[6,7] and other extracellular proteins,[8] owing to their high intrinsic flexibility. The presence of conformational plasticity, or "shape-shifting," becomes relevant for homotypic and heterotypic complex formation of Contactins and the selection of specific binding sites by various molecules. However, conventional structure determination methods provide a technique-based snapshot of flexible, multi-domain biomolecules, falling short in defining their structural flexibility and dynamics. X-ray diffraction typically biases toward a single conformation in the crystal lattice, while single-particle cryo-EM reconstructions offer a single view or limited view of conformational variability, averaging out flexible portions in both techniques. NMR (Nuclear Magnetic Resonance), the best approach to probe and measure protein flexibility, is limited by protein size. Nonetheless, observed conformations differ depending on the methods used in structural studies, suggesting high structural flexibility within the domain conglomerate of the Contactin family.

Human and mouse Contactin-2 are heavily glycosylated, containing eleven N-linked sites, creating vast structural heterogeneity depending on the types, lengths, and linkage of the carbohydrate components. For instance, Contactin-1 glycan composition regulates its interactions with NF155 and endows specificity for autoantibodies in patients with chronic inflammatory demyelinating polyradiculoneuropathy.[9] Beyond intrinsic structural

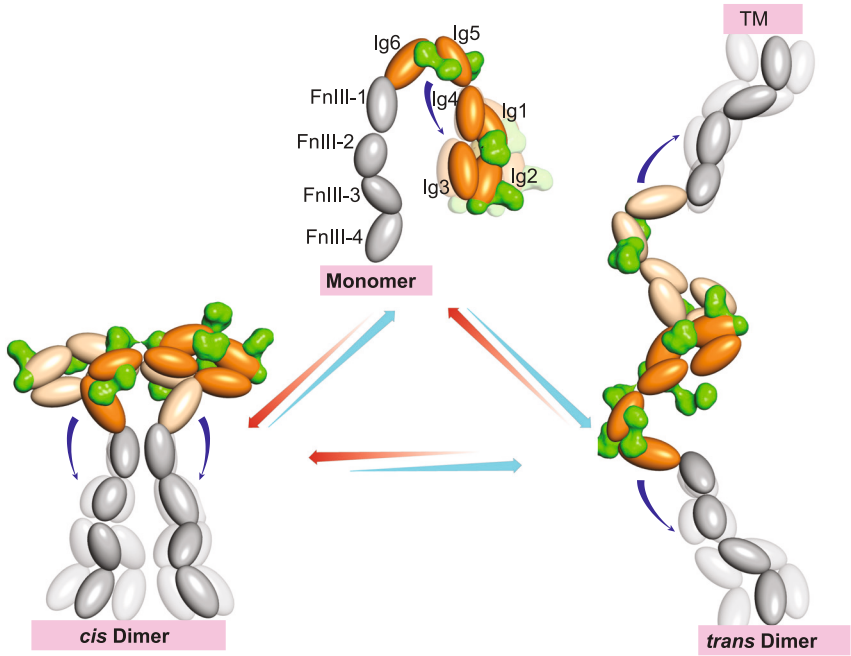

Figure 1. Representative Contacin-2 dimeric and monomeric states in conformational equilibria

The monomeric Contactin-2 is compact adopting a bent-closed conformation, whereas its dimeric form shows homotypic interaction with one another in two, if not more, states *in cis* and *trans*. Glycans were added at Contactin-2 N-glycosylation sequons using the Contacin-2 crystal structure.[3] Intersubunit crowding and electrostatic repulsions due to sugar coating(s) may regulate *cis* and *trans* homotypic interactions as well as expose or shield epitopes for the heterotypic binding in such adhesion molecules analogously to the regulation of conformational transitions in integrins and CAM molecules. Individual domains from PDB# 8AOY are shown as 3D ellipses, and N-glycans are shown as green molecular surfaces. Images were prepared using PyMOL.

flexibility, the "sugar coating" of the Contactin family also emerges as biologically significant, adding an extra layer of complexity to their structural plasticity. Here, the authors pose more questions than answers about conformational equilibria and allosteric regulation, crucial for regulating homotypic and heterotypic interactions of Contactin-2 as well as the structural and functional features of many other extracellular molecules. Contactin-2 belongs to a large class of highly glycosylated molecules with extracellular domains in tandem, many of which are multi-domain and often cysteine-rich. The emerging hypothesis is that conformational dynamism often regulates the functional output of these proteins through a graded or rheostat-like mechanism.[10] Changes in expression level, glycosylation, phosphorylation, or any other post-translational modifications (PTMs) could be subtle inputs for the temporal and spatial regulation of the conformational equilibrium (Figure 1) that directs specific biological processes (homotypic and heterotypic protein interactions) and, when altered, could give rise to a variety of pathologies.

Chataigner et al., using in-solution observables, provided evidence for the roles of N-linked sugars and their coating composition in Contactin-2 homo-dimerization. Mannosylated glycans favored a higher-order oligomeric state while the complex sugar coating showed the opposite effect. Alanine mutations of two highly conserved N-glycan-bearing Asn residues clearly elevated the dimer level, further supporting the roles of sugars and composition of sugar coatings in Contactin-2 homotypic interaction. The authors also provided mechanistic insight into the overlooked massive conformational space that N glycans occupy on the surface of these molecules. Glycosidic bonds characteristically have free-rotation at the Asn-monosaccharide and monosaccharide-monosaccharide linkages, imparting them, if not fully but partially, difficult to visualize in the electron density map. This intrinsic flexibility, along with terminal charged sialic acid residues, also endows each glycan to sweep a large hydrodynamic volume over a domain it is attached to, creating both "crowding" and electrostatic repelling effects. These findings reinforce the principle that variation in N glycan number and composition regulate structural plasticity, and thus, conformational equilibria of multi-domain cell surface molecules. Conformational flexibility and equilibria of extracellular proteins and their glycosylation events should continue to be an interesting area to explore and will likely provide fascinating insights into the biological events, especially those associated with human health and pathologies.

ACKNOWLEDGMENTS

M.S. receives an NSF CAREER award (2239492). I want to thank Zacchaeus O. Alabi and Sean-Patrick Scott for their support in preparing this preview.

DECLARATION OF INTERESTS

The author declares no competing interests.

REFERENCES

1. Lu, Z., Lei, D., Seshadrinathan, S., Szwed, A., Liu, J., Liu, J., Rudenko, G., and Ren, G. (2018). 3D Images of Neuronal Adhesion Molecule Contactin-2 Reveal an Unanticipated Two-State Architecture. bioRxiv, 386102.

2. Freigang, J., Proba, K., Leder, L., Diederichs, K., Sonderegger, P., and Welte, W. (2000). The crystal structure of the ligand binding module of axonin-1/TAG-1 suggests a zipper mechanism for neural cell adhesion. Cell *101*, 425–433.

3. Chataigner, L.M., Thärichen, L., Beugelink, J.W., Granneman, J.C., Mokiem, N.J., Snijder, J., Förster, F., and Janssen, B.J. (2023). Contactin 2 homophilic adhesion structure and conformational plasticity. Structure, 60–73.

4. Lu, Z., Reddy, M.V.V.V.S., Liu, J., Kalichava, A., Liu, J., Zhang, L., Chen, F., Wang, Y., Holthauzen, L.M.F., White, M.A., et al. (2016). Molecular Architecture of Contactin-associated Protein-like 2 (CNTNAP2) and Its Interaction with Contactin 2 (CNTN2). J. Biol. Chem. *291*, 24133–24147.

5. Schürmann, G., Haspel, J., Grumet, M., and Erickson, H.P. (2001). Cell adhesion molecule L1 in folded (horseshoe) and extended conformations. Mol. Biol. Cell *12*, 1765–1773.

6. Jun, C.D., Carman, C.V., Redick, S.D., Shimaoka, M., Erickson, H.P., and Springer, T.A. (2001). Ultrastructure and function of dimeric, soluble intercellular adhesion molecule-1 (ICAM-1). J. Biol. Chem. *276*, 29019–29027.

7. Meijers, R., Puettmann-Holgado, R., Skiniotis, G., Liu, J.H., Walz, T., Wang, J.H., and Schmucker, D. (2007). Structural basis of

Dscam isoform specificity. Nature *449*, 487–491.

8. Pinelo, J.E.E., Manandhar, P., Popovic, G., Ray, K., Tasdelen, M.F., Nguyen, Q., Iavarone, A.T., Offenbacher, A.R., Hudson, N.E., and Sen, M. (2023). Systematic mapping of the conformational landscape and dynamism of soluble fibrinogen. J. Thromb. Haemost. *21*, 1529–1543.

9. Labasque, M., Hivert, B., Nogales-Gadea, G., Querol, L., Illa, I., and Faivre-Sarrailh, C. (2014). Specific contactin N-glycans are implicated in neurofascin binding and autoimmune targeting in peripheral neuropathies. J. Biol. Chem. *289*, 7907–7918.

10. Li, J., Su, Y., Xia, W., Qin, Y., Humphries, M.J., Vestweber, D., Cabañas, C., Lu, C., and Springer, T.A. (2017). Conformational equilibria and intrinsic affinities define integrin activation. EMBO J. *36*, 629–645.

From simple to complex: Reconstructing all-atom structures from coarse-grained models using cg2all

Yui Tik Pang,[1] Lixinhao Yang,[2] and James C. Gumbart[1,2,*]
[1]School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA
[2]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA
*Correspondence: gumbart@physics.gatech.edu
https://doi.org/10.1016/j.str.2023.12.004

In this issue of *Structure*, Heo and Feig present cg2all, a novel deep-learning model capable of efficiently predicting all-atom protein structures from coarse-grained (CG) representations. The model maintains high accuracy, even when the CG model is simplified to a single bead per residue, and has a number of promising applications.

Deep-learning methodologies have undergone swift advances since the introduction of the transformer architecture in 2017.[1] Transformers leverage a deep-learning technique known as "attention," which allows neural-network models to focus on interrelated segments within a data sequence when generating outputs. This approach has demonstrated exceptional efficacy in a wide range of generative deep-learning applications, from natural language processing to computer vision, yielding highly realistic and convincing results. In the realm of biology, AlphaFold2 employs transformers to discern patterns in protein sequence data, achieving unparalleled accuracy in protein structure prediction.[2] In many cases, these predictions are precise matches to experimental structures.

However, a machine learning method that can predict protein conformations beyond just one (or a few) structures is still missing, and molecular dynamics (MD) simulation remains the predominant tool to study protein dynamics in silico. Traditionally, MD simulations represent proteins and their surrounding biological environments atom by atom, which contributes to their accuracy but also incurs a high computational cost. One approach to mitigate the computational cost of MD simulations is the use of coarse-grained (CG) models, which reduce the resolution of the system from atomistic to several beads per residue. Over the years, CG models have been utilized to address numerous questions across fields such as biology and materials science.[3]

Coarse graining enables MD simulations to explore protein conformations more rapidly; however, the reduced resolution limits the accuracy and insights that can be derived. For example, hydrogen bonds and salt bridges play important roles in protein structure and dynamics; yet, without hydrogen atoms, CG models fail to capture these interactions explicitly. Prior attempts using library-based methods to reconstruct all-atom models from CG models[4,5] have achieved modest accuracy but frequently require extensive optimization to eliminate steric clashes and other physical imperfections.

Here, Heo and Feig introduce cg2all, a deep-learning approach that converts CG models back into their corresponding all-atom representations (Figure 1).[6] Cg2all utilizes the SE(3) transformer,[7] a variant of the self-attention mechanism that remains equivariant under 3D roto-translations, as well as a rigid-body block representation of the protein inspired by AlphaFold2, to construct all-atom structures from CG models at various resolutions. The model is trained using high-resolution X-ray crystal structures from the Protein Data Bank (PDB), with a loss function that incorporates both data-dependent terms as well as physics-based terms, such as torsion energies.

The authors trained individual instances of cg2all for a variety of CG models, with resolutions ranging from one to eight beads per residue, including some well-established models such as MARTINI[8] and PRIMO.[9] All models exhibited excellent reconstruction accuracy for the validation set, with the average heavy-atom RMSD being 0.31 Å and 0.18 Å for reconstructions from MARTINI and PRIMO, respectively. Remarkably, cg2all also achieved an average heavy-atom RMSD of 0.46 Å when reconstructing from the center-of-mass one-bead-per-residue model, underscoring cg2all's capability to predict and construct side-chain-atom coordinates using only