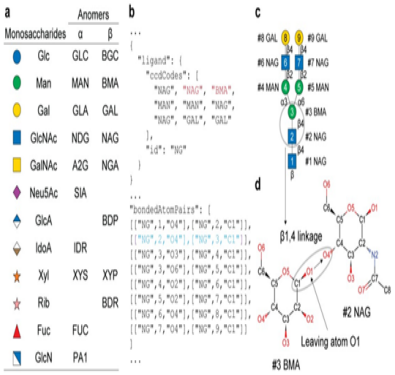

Modeling glycans with AlphaFold 3: capabilities, caveats, and limitations

Description

Glycans are complex carbohydrates that exhibit extraordinary structural complexity and stereochemical diversity while playing essential roles in many biological processes, including immune regulation, pathogen recognition, and cell communication. In humans, more than half of all proteins are glycosylated, particularly those in secretory and membrane-associated pathways, highlighting the importance of glycans in health and disease. The recent release of the AlphaFold 3 source code enables customizable modeling not only of proteins but also glycan-containing biomolecular complexes. The authors assessed the capacity of AlphaFold 3 to model glycans using several input formats and identified a hybrid syntax employing Chemical Component Dictionary (CCD)-based molecular building blocks linked by â??*bondedAtomPairsâ??* (*BAP*) as most effective in generating stereochemically valid glycan models. This workflow was used to create a library of AlphaFold 3 input templates and corresponding structural models for various glycan classes. Further exploration of capabilities, limitations, and remediation strategies for modeling problematic structures was conducted. Glycan interactions were also modeled with glycosylation enzymes and lectins, with benchmarking and validation against known crystal structures. This protocol-driven approach is valuable for generating stereochemically valid, static models of glycan-protein interactions to support hypothesis development and subsequent structural and functional validation. However, caution should be observed in the overinterpretation of the static models since glycans are known to exhibit considerable conformational dynamics that can be further captured by equilibrium sampling using molecular dynamics-based approaches. By sharing benchmarked examples using the BAP syntax the authors aim to support broader evaluation of AlphaFold 3 in studying glycan- related mechanisms in biosynthesis, signaling, infection, and disease.

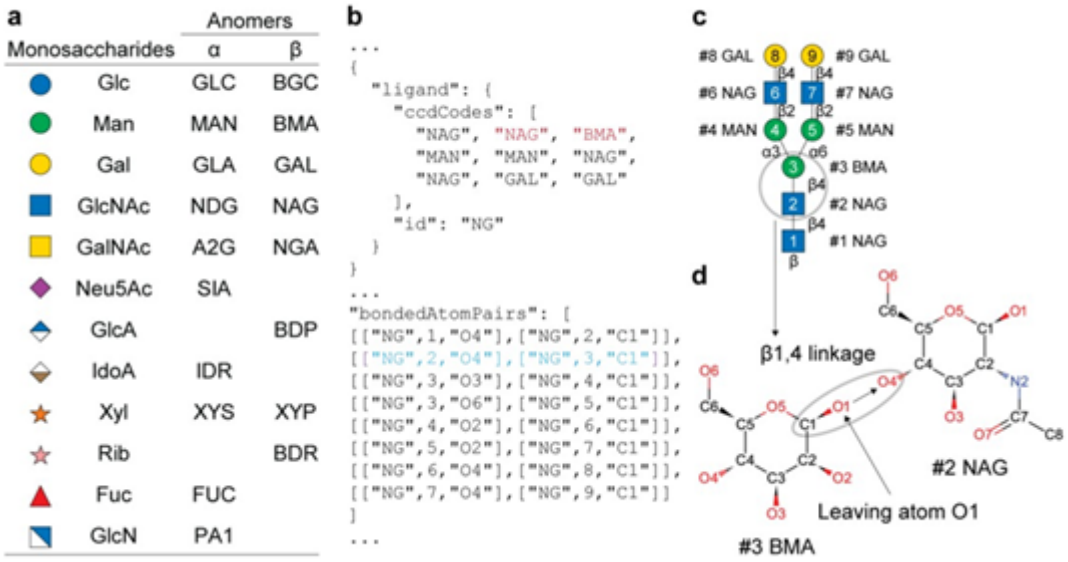

Figure. Glycans specification using Chemical Component Dictionary (CCD) and *bondedAtomPairs* (*BAP*) syntax in AlphaFold 3**. a,** Representative CCD codes for common human monosaccharides. **b,** Partial JSON script used to generate G2 *N*-linked glycan. **c**, SNFG representation of G2 *N*-linked glycan with residue numbering equivalent to what is used in the JSON modeling input file. The order of the monosaccharides in the JSON file *ccdCodes* list corresponds with the residue numbering in the SNFG representation. **d**, Demonstration of a Î²1,4 linkage between GlcNAc and Man, specified by the second and third CCD entries (red NAG and BMA) in panel **b**. The glycosidic bond is defined by the *bondedAtomPairs* (*BAP*) field (blue in panel **b**), specifying a bond between the O4 atom of NAG and the C1 atom of BMA (gray oval in panel **d**). The entry in the *bondedAtomPairs* (*BAP*) list leads to covalent bond formation between the respective hydroxyl oxygen (O4 of NAG) and the anomeric C1 (C1 of BMA) with the removal of the Î²-linked O1 (leaving atom) on the C1 of BMA. The resulting glycosidic linkage retains the anomeric configuration at C1 of the original CCD entry (Î²-linkage for BMA).

Category

1. News