Electronic Supplementary Material

A topology framework for macromolecular complexes and condensates

Maziar Heidari[1], Duane Moes[1], Otto Schullian[1,2], Barbara Scalvini[1], and Alireza Mashaghi[1,3] (✉)

[1] *Medical Systems Biophysics and Bioengineering, Leiden Academic Centre for Drug Research, Faculty of Science, Leiden University, Leiden 2333CC, the Netherlands*
[2] *Fachbereich Physik, Freie Universität Berlin, Arnimallee 14, Berlin 14195, Germany*
[3] *Harvard Medical School, Harvard University, 25 Shattuck St, Boston MA 02115, USA*

Supporting information to https://doi.org/10.1007/s12274-022-4355-x

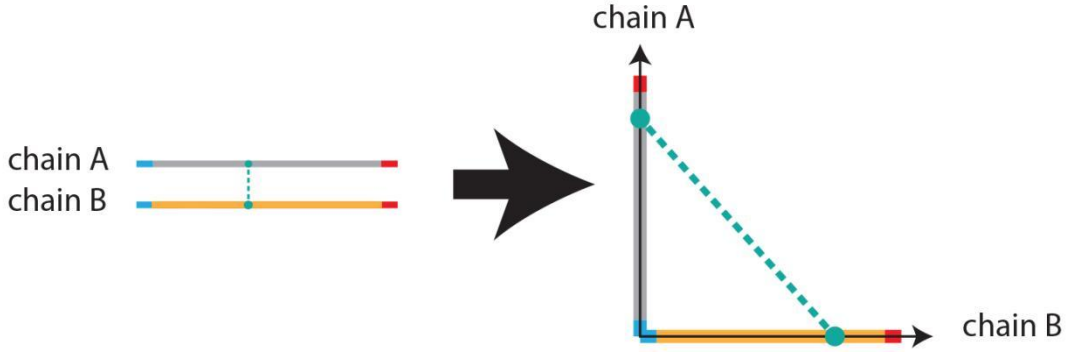

Figure S1 Procedure to represent a contact in an n-dimensional vector space. The chains are placed on separate axes and the contact is represented by a straight line between the two contact sites.

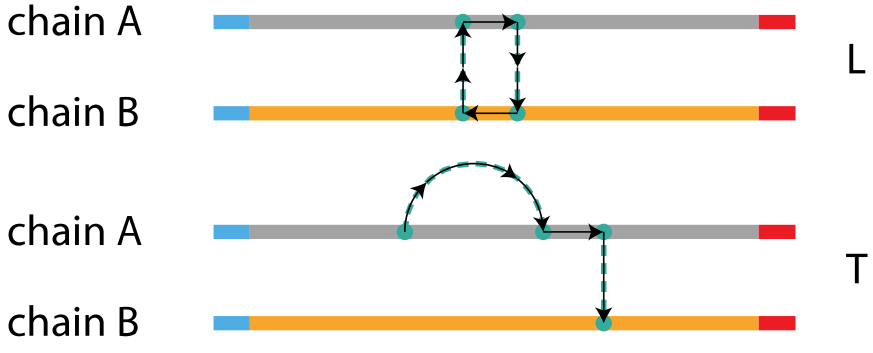

Figure S2 Procedure to determine the loop(L)-configuration (top) and the tandem(T)-configuration. If two contacts form a closed loop including segments of the chains, without going through the same path twice, then the system has L-configuration. If both contacts can be joined a path but do not form a loop, the system is in T-configuration. In all other cases the system is in I-configuration.

Address correspondence to a.mashaghi.tabari@lacdr.leidenuniv.nl

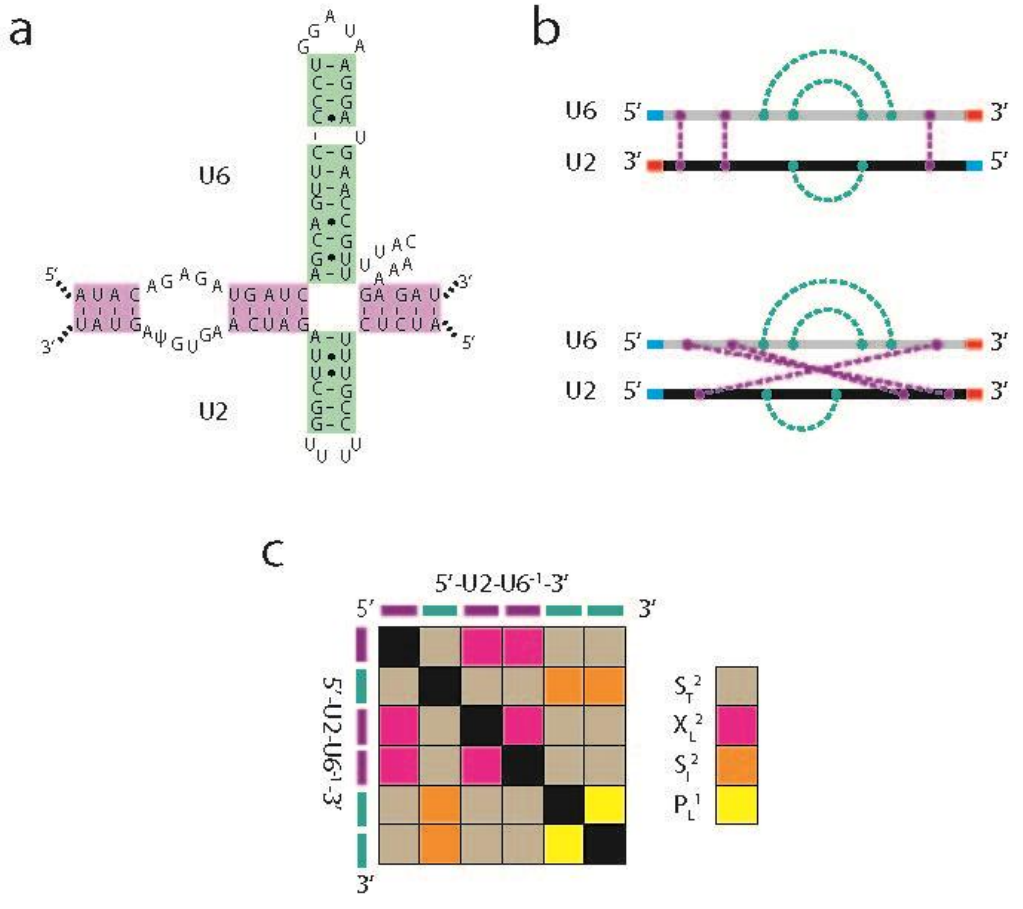

Figure S3 Structure and topology analysis of the spliceosomal U2-U6 complex, with contacts formed by base-pairing of nucleobases. a: Structure with the identified contacts (green and purple) b: Sketch of the contacts with straight chains without (top) and with (bottom) taking the direction into account. c: Matrix representing the relations between all contacts.

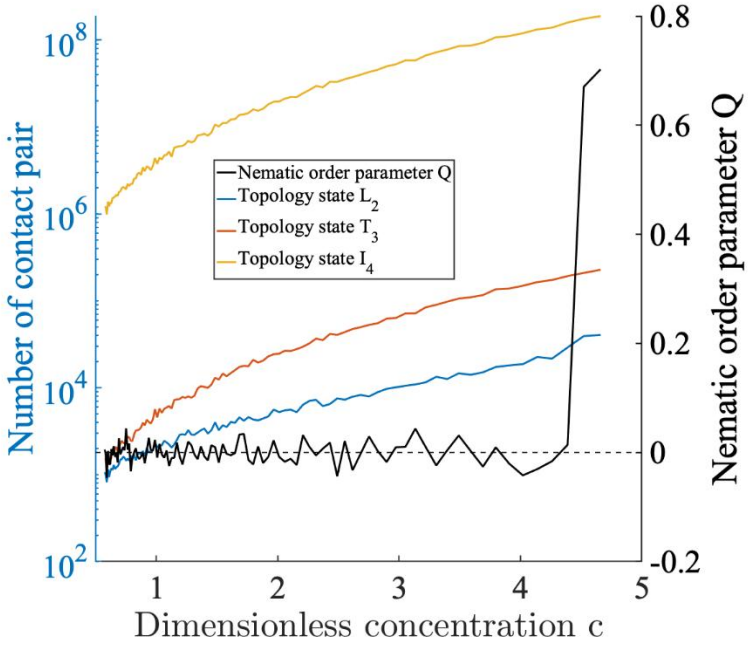

Figure S4 The topology fraction L_2, T_3 and I_4 and nematic order parameter Q as a function of dimensionless concentration c. The topology fraction L_2 only exhibits discontinuity at the concentration of isotropic-nematic transition.

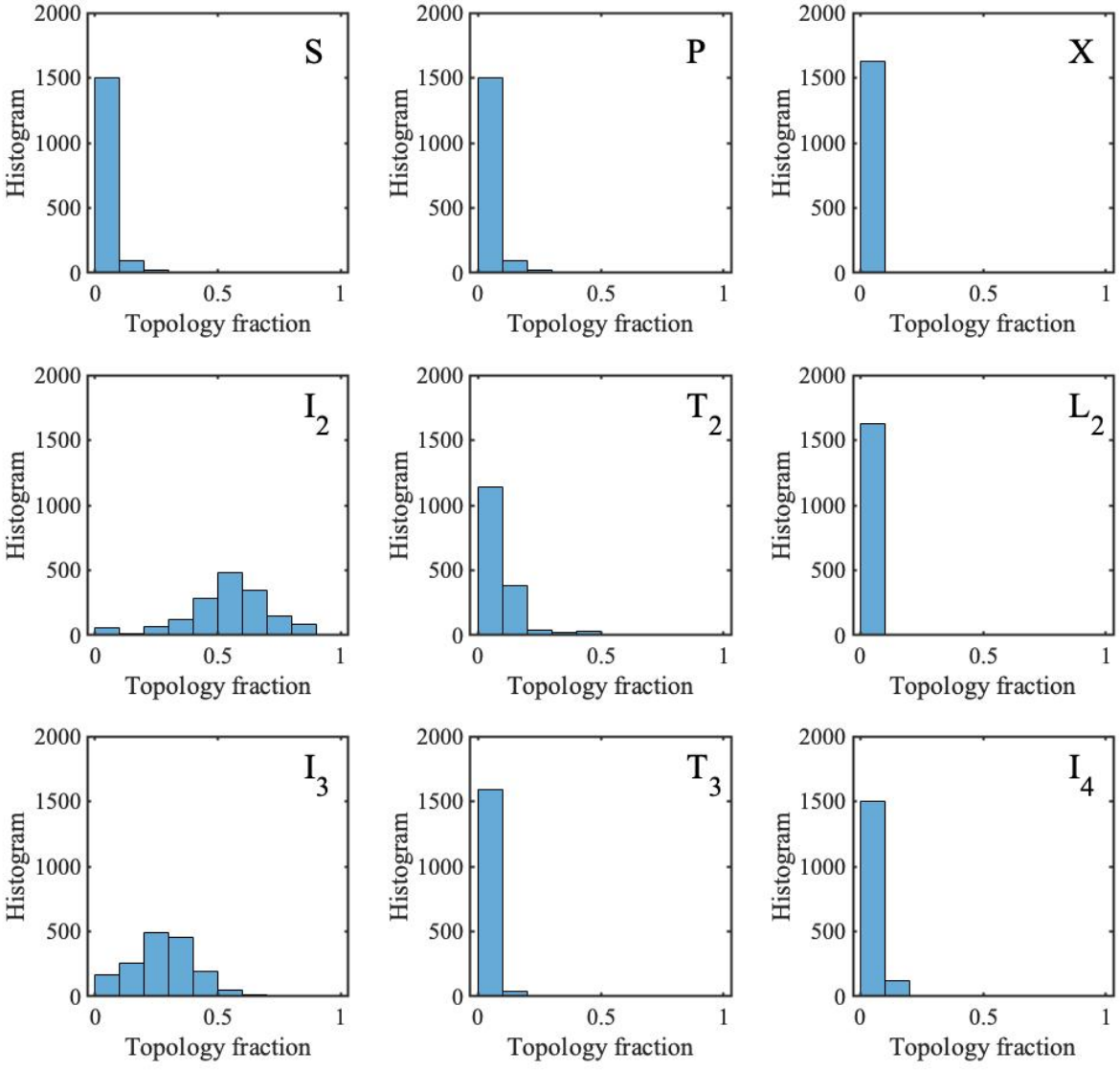

Figure S5 Circuit topology of protein complexes. The distribution of the topology fractions of protein complexes. The number of protein complex is 170.

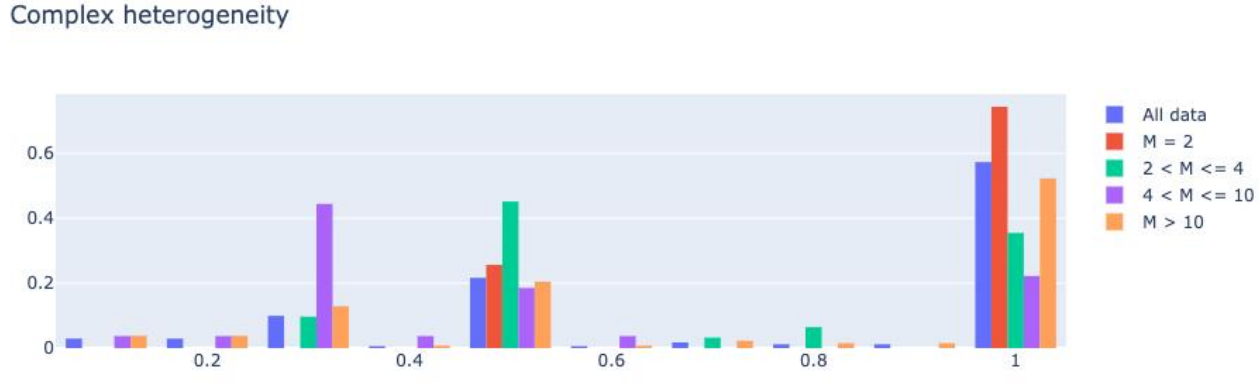

Figure S6 Heterogeneity index of protein complexes as categorized by the size of the complex (M_c). The number of protein complex is 170.

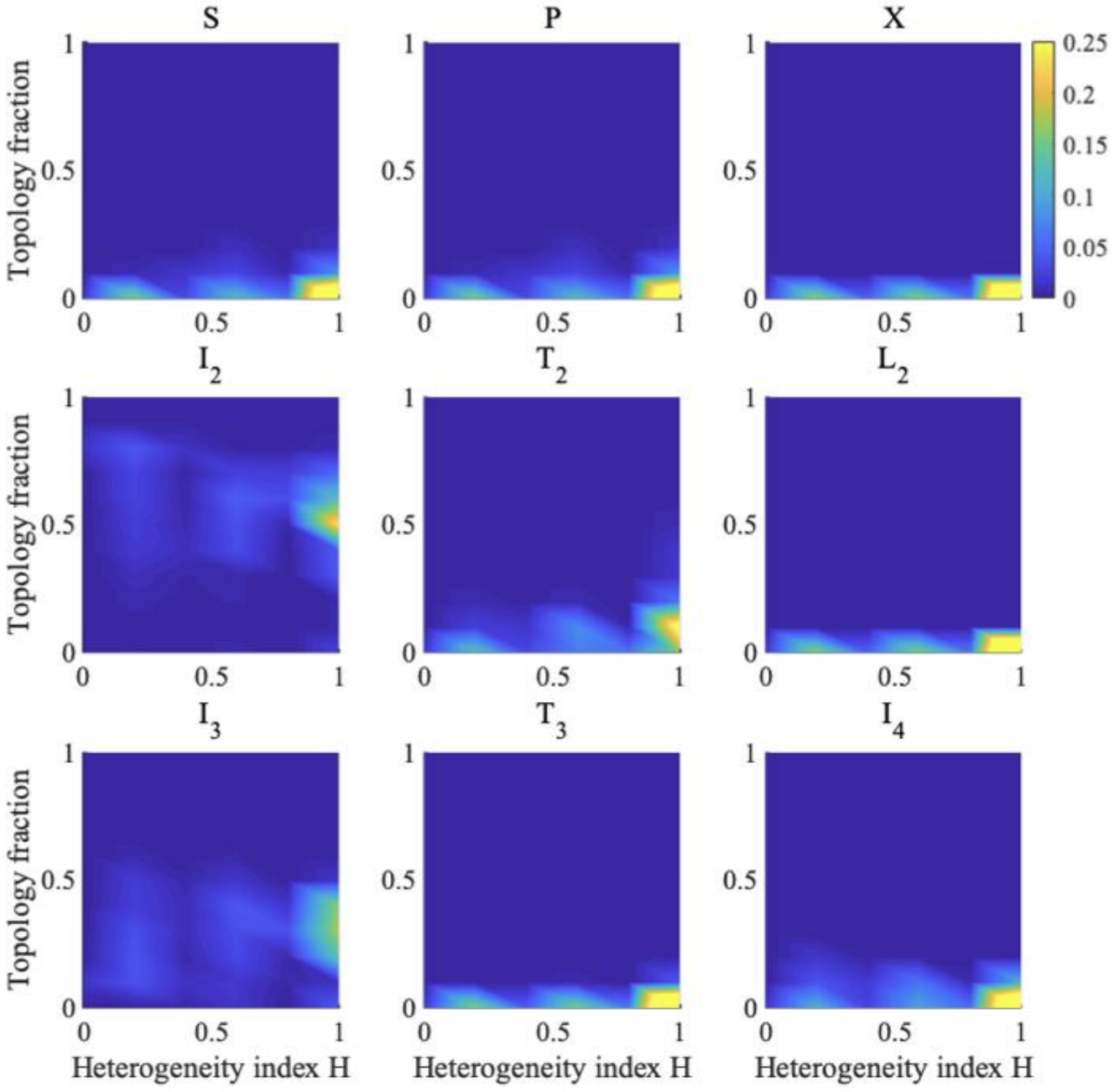

Figure S7 Circuit topology of protein complexes. Bivariate plots of topology fractions of protein complexes and the complex heterogeneity index (H).

Appendix: Other protein complex topology frameworks involving contacts

Network analysis is an alternative approach to analyzing non-covalent contacts within folded single polymer chains (Di Paola et al., 2013, Holm and Sander, 1996, Kladwang et al., 2011). A network representation of intramolecular interactions captures connectedness and modularity of structures and allows the use of a plethora of graph theoretical methods to find valuable properties such as shortest path, cluster size and the existence of hubs (Albert and Barabási, 2002, Mashaghi et al., 2004, Newman, 2003, Ramezanpour et al., 2003). The relationship between the topology of protein contact networks (PCN) and protein function and dynamics is well established. For example, shortest

paths in a PCN appear to mediate concerted motions and energy transmission upon ligand binding (Bastolla et al., 2013, Berezovsky, 2013). Comparative analysis of PCN associated with the folding transition state and the native state reveals critical contacts that drive nucleation and subsequent growth to the native state (Vendruscolo et al., 2001). Network analysis can also be used to study complexes. One can represent the chains as one tandem chain and build a PCN similar to single molecules (Di Paola and Giuliani, 2015). Here, the order and orientation of chains affects the contact map. Alternatively, every chain can be represented as a node and contacts as links between the nodes. A number of properties can be readily extracted from this representation, such as degree distribution, clustering coefficients or link weight statistics.

References:

Albert, R. and Barabási, A.-L. (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* 74, 47-97.

Bastolla, U., Porto, M., and Roman, H.E. (2013) The emerging dynamic view of proteins: protein plasticity in allostery, evolution and self-assembly. *Biochimica et biophysica acta* 1834, 817-819.

Berezovsky, I.N. (2013) Thermodynamics of allostery paves a way to allosteric drugs. *Biochimica et biophysica acta* 1834, 830-835.

Di Paola, L. and Giuliani, A. (2015) Protein contact network topology: a natural language for allostery. *Current opinion in structural biology* 31, 43-48.

Holm, L. and Sander, C. (1996) Mapping the protein universe. *Science* 273, 595-603.

Kladwang, W., VanLang, C.C., Cordero, P., and Das, R. (2011) A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nature chemistry* 3, 954-962.

Mashaghi, A., van Wijk, R.J., and Tans, S.J. (2014) Circuit Topology of Proteins and Nucleic Acids. *Structure* 22, 1227-1237.

Newman, M.E.J. (2003) The Structure and Function of Complex Networks. *SIAM Review* 45, 167-256.

Ramezanpour, A., Karimipour, V., and Mashaghi, A. (2003) Generating correlated networks from uncorrelated ones. *Physical review. E, Statistical, nonlinear, and soft matter physics* 67, 046107.

Vendruscolo, M., Paci, E., Dobson, C.M., and Karplus, M. (2001) Three key residues form a critical contact network in a protein folding transition state. *Nature* 409, 641-645.