First let us make some short introduction and have a look at the methods used in the **KnotGenome server**: https://knotgenom.cent.uw.edu.pl. We will browse **Read More** page:

1. . ***Introduction*** *section, bottom:*

   The data that all the examples in the KnotGenome server are based on were provided by:

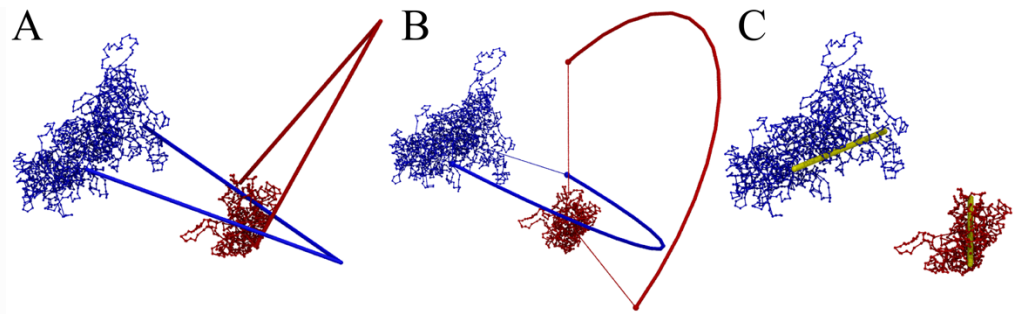
   Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O'Shaughnessy-Kirwan, A, et al. Nature 2017, 544, 59—64.

   3D structures of whole genomes of single haploid mouse embryonic based on Hi-C chromosome conformation contact data, from 8 cells.
   Each monomer corresponds to less than 100kb.

2. . ***Knot detection and classification*** *section:*

   Knot types are determined by computing polynomial knot invariants, in this case **Alexander polynomial and Homfly-PT polynomial**.
   To calculate these invariants first the chain has to be closed, and on KnotGenome we propose 3 methods:
   - (C) direct segment connecting two ends;
   - (B) the centre of mass method, where the chromosome endpoints are connected to two points on the sphere based on directions determined by their positions and the centre of mass of structure, and these two points are connected by an arc lying on the surface of the sphere;
   - (~A) finally random closure method, where number of times we randomly choose two points on the big sphere, connect endpoints to these points and then connect these two points by an arc lying on the surface of the sphere; for each such random closure knot type is determined and the one occurring most frequently is choosen as a knot type of oryginal chain.

   The centre of mass method was not available on the Knotprot and Linkprot servers. It is way faster than statistical method and worth considering in some long computations, especially for single chains.

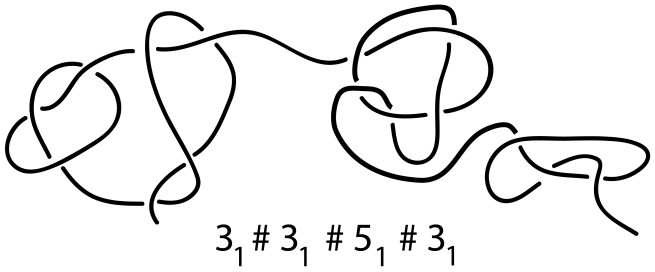


3. . ***Link detection and classification*** *section:*

   In the case of links we use only **Homfly-PT polynomial** (since Alexander polynomial doesn't differ links so well). Here we have 3 available methods to close chains as well, the only difference is in the statistical method where for each chain we choose one random point on the sphere. Two points would be similar to the centre of mass method where is a high risk that those wide added arms will be linked and will introduce artificial linking between two chains.

4. The algorithm used by the KnotGenome distinguishes knots and links up to (besides knot fingerprint matrix):

	Prime knots	*Composite knots*	*Prime links*	*Composite links*
Up to # crossings	10	12	14	~8
Example	3_1	$3_1 \# 4_1$	Hopf	3_1#4_1#Hopf
Type on the KnotGenome	3_1	$3_1 \# 4_1$	Hopf	**Other**

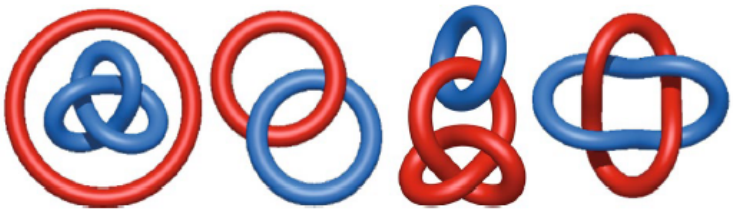


An example of a complex composite knot found along the chromosome chain using KnotGenome *(two 3_1 prime knots on the left are not placed separable along the chain).*

A prime knot is a knot which cannot be decomposed into a sum of two or more simpler knots. **Composite knots** are knot sums of prime knots, for instance $3_1 \# 3_1$ denotes a knot made of two trefoil knots. In the projection of a composite knot one can draw a line which divides it into two parts, each of which contains an independent knot.

5. . ***GLN calculations*** *section*

There is a lot of composite knots and links in genome data (in opposite to proteins described yesterday), and especially in the case of links – if chromosomes (chains) are entangled, we obtain often "Other" type which does not tell us a lot.
This is why we use Gauss Linking Number in the case of links as well. GLN indicates how many times one curve winds around the second one.



Unknot U 3_1 Unlink GLN = 0 | Hopf Link GLN = +/- 1 | 3_1 # Hopf Link GLN = +/- 1 | Solomon Link GLN = +/- 2

We can calculate GLN for open chains as well, according to the formulas given in *GLN calculation* section of *Read more*. Value around 0 tells us that chains are unlinked, around +-1 that they are linked in Hopf style and so on. *(whGLN, max|GLN|, max short |GLN| will be explained while looking at the examples)*.

6. .***Relaxation procedure*** *section*
   Molecular dynamics simulations of the 3D genome were introduced to assess the significance of entanglements present in the experimentally derived nuclear chromosomal structures. We use a structure based representation of the system, in which the provided structure minimizes the potential function. Adjacent beads along the chain interact with spring potential, and angle, and dihedral potentials. If two non-adjacent beads along the chain are in a distance smaller than 0.6 nm, they do not interact with each other (there is no penalty for overlapping); if they are within a range of 0.6 - 2.0 nm in the uploaded structure then they interact with the Lennard-Jones potential; otherwise they only repulse each other at the short distances.

   After a simulation run, 10 representative configurations are analyzed in order to observe the evolution of entanglements in the studied system and assess whether it is stable or not.

   The details of the relaxation method and procedure are presented here:
   https://knotgenom.cent.uw.edu.pl/relax_procedure

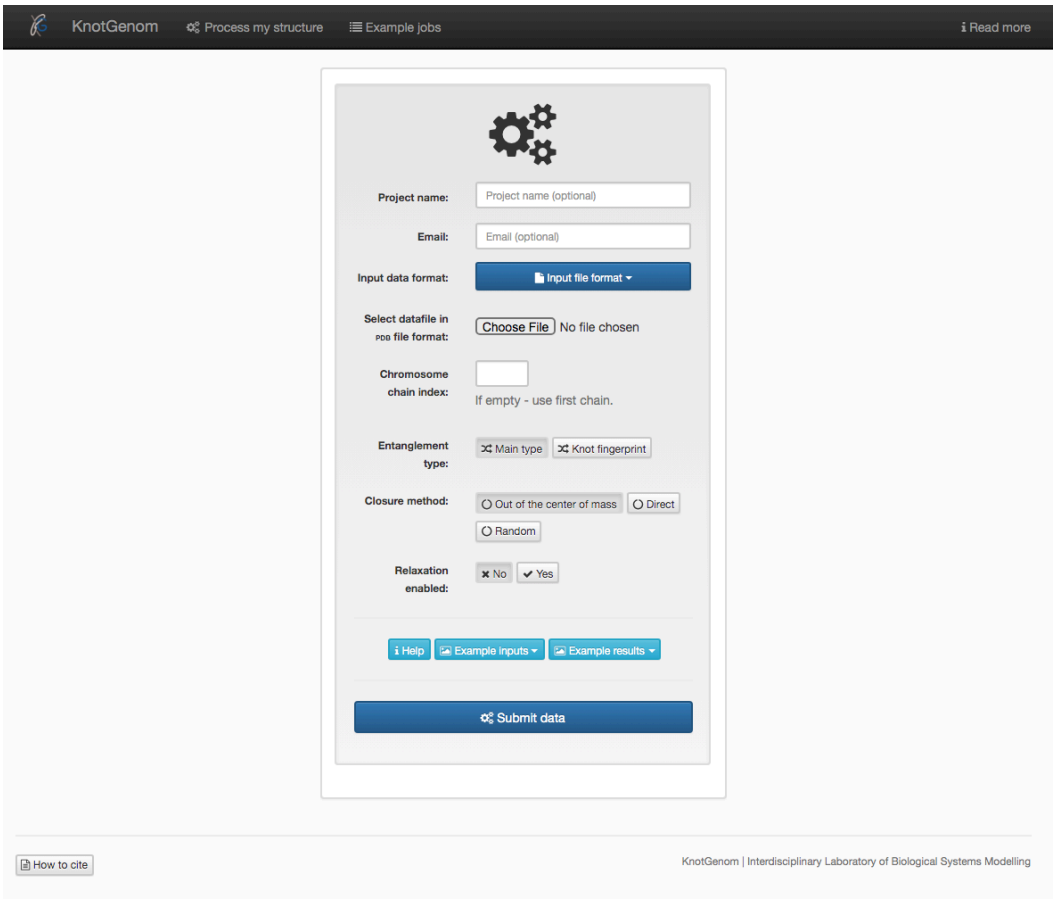
Now lets test the **server**. Click **Process my structure"** button in the top menu or at the bottom of the page. It will direct you to the **submit menu**.

7. First two fields are optional: **Project name** and **e-mail address**. Providing "**Project name**" would be helpful in the case of running several jobs to distinguish between them. If the e-mail address is provided, a link to the results of the job will be send to this address after the job is submitted.

   Before uploading the structure for the job, one has to determine an **input data format**. The server accepts three formats: **PDB**, **XYZ** and **N3D**, with single chain or "whole cell" with 2 up to 40 chains in one file. For details of all formats one can click this light blue "**iHelp**" button.

   On the github (https://github.com/ilbsm/databases_api_tutorial) you have few files with single chromosomes and pairs of chromosomes from our data from mouse cell nr 5 (in XYZ and PDB formats).

8. Once "**Input file format**" is set, one can upload a structure file for the calculations on the server. Let's consider the PDB file format with single chromosome option first. There appear several new fields: (1) **select input structure**. You can drag-and-drop the file, or browse it, (2) choose chain name, (3) one can decide whether to calculate just the main knot type of a structure or whole knot fingerprint matrix, (4) choose the "**Closure method**" (there are three options discussed before, "Out of the center of mass", "Direct" (connect both ends with straight line) and "Random" (here we can (5) determine the number of random closures used for the statistics), and (6) decide whether **relaxation** (short molecular dynamics) should be performed.



9. Let's start with chromosome `h` (`05_01_h.pdb` file). In the file there is only chain h, thus there is no need to choose chain index. We will calculate just the knot type of whole structure and for now let's leave default closure method and do not perform relaxation.

   If we haven't provided an email address, the only way to reach the result page is to follow the link in the green box. Here we can observe **the status of a job**. Usually every one minute jobs are loaded for computations, but you may need to wait even 5 minutes. In another tab, and we can run the next job. Lets try different closure methods. We can compare results with the calculated earlier jobs:

   - **[EXAMPLE-1: PDB, `05_01_h.pdb`, main type, out of center ->** $3_1\#4_1$**]**,

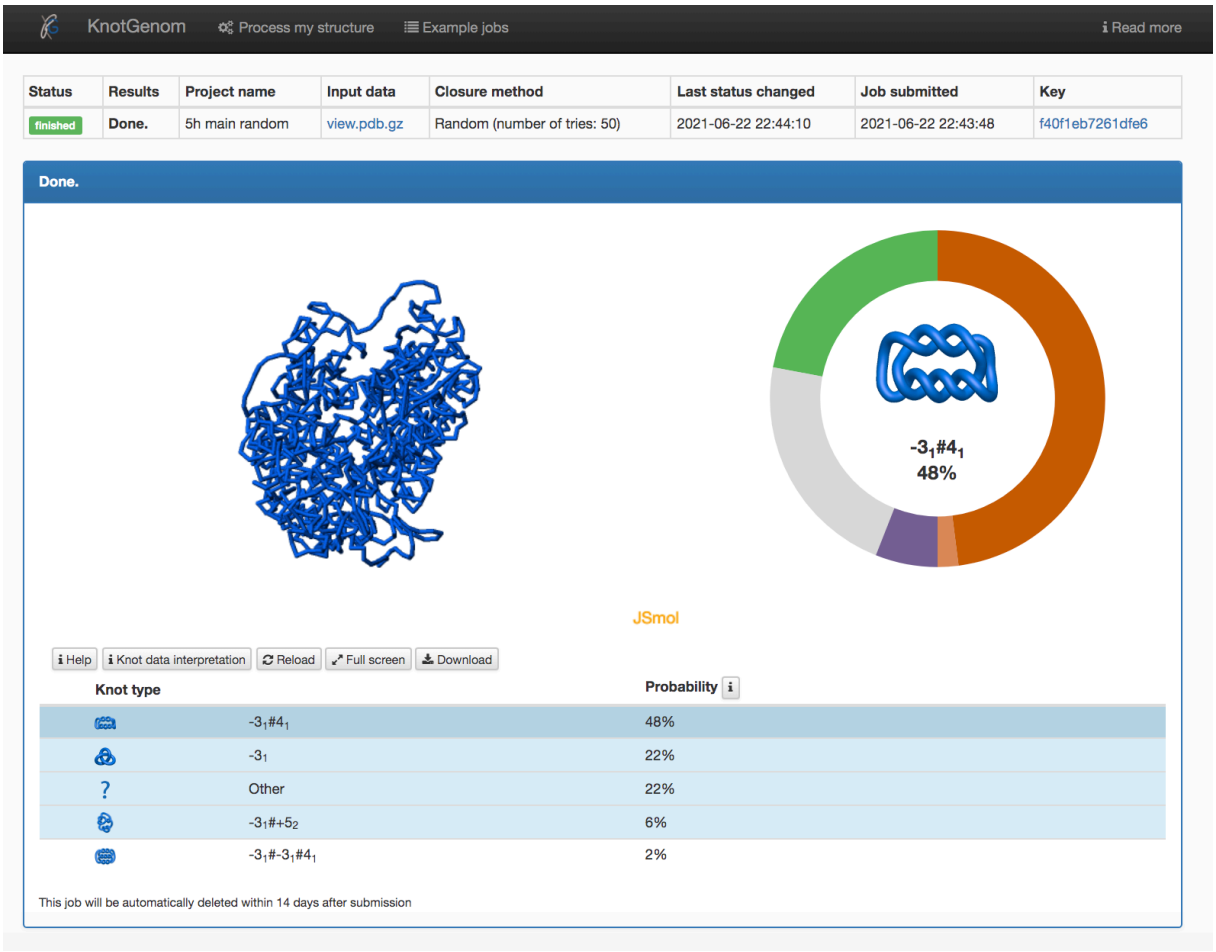
     results: https://knotgenom.cent.uw.edu.pl/compute/ed2f485dc30a61a,

   - **[EXAMPLE-2: XYZ, `05_01_h.pdb`, main type, direct method->** $3_1\#3_1\#4_1$**]**,

     results https://knotgenom.cent.uw.edu.pl/compute/2271cc5f9bdd012,

   - **[EXAMPLE-3: XYZ, `05_01_h.pdb`, main type, random method ->** 48% of $3_1\#4_1$**]**,

     results, https://knotgenom.cent.uw.edu.pl/compute/f40f1eb7261dfe6.



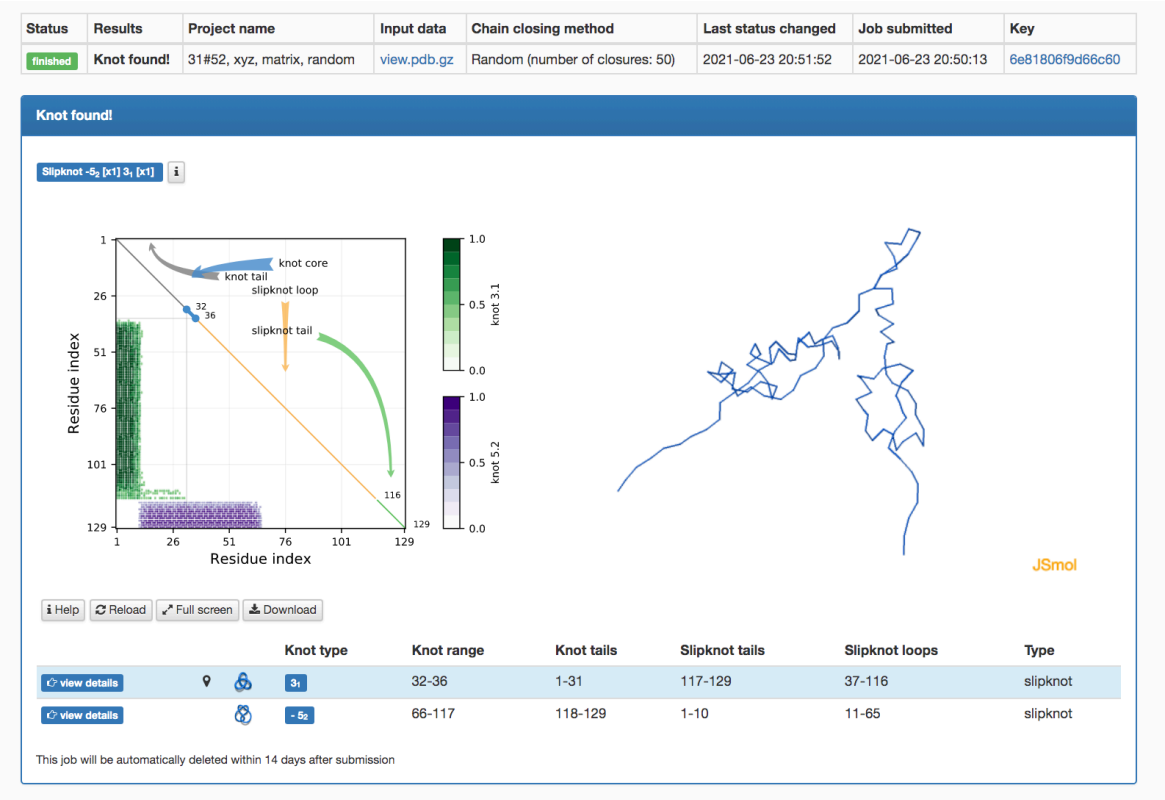
One can notice that most probable knot type, 31#41, has appeared with "Out of the center of mass" method, while with "Direct" method we've got even more complex knot. The last is quite common when ends of chain are not located very close to each other, and added segments with high probability can complicate entanglement. Thus for faster computations (for single chains) we recommend rather "Out of the center of mass" method.

Always you can download several files, in this case the structure in two formats and results.

10. You can also compute whole knot fingerprint matrix, very similar to the KnotProt server main function. This is way more time consuming job, for chromosomes of length around 1000 it would take from few minutes to many hours, depends on the settings. Lets try with shorter structure, `knots-31-52.xyz` file, use "Random" closure method with 50 tries. Blue button "Show advanced" has appeared – now we can choose the resolution of the knot fingerprint matrix – by default only points on the matrix on the net of length 7 will be checked, and if the knot occurs with frequency higher than 20% in the point the neighborhood of that particular point on the matrix will be checked as well. This is a nice option that wasn't available at the Knotprot, it may make computations faster. But you need to be careful using it.

    - **[EXAMPLE-4: XYZ, `knots-31-52.xyz`, knot fingerprint, random method** (50)],

results, https://knotgenom.cent.uw.edu.pl/compute/6e81806f9d66c60



Here we have two knots in the structure. Here we should have different color for 31#52 knot, but KnotGenome recognizes it as Other (it does not has appropriate polynomial in its database). You can download the maps and file with raw data.

11. Pair of chromosomes r and t (input file format: whole cell, XYZ):

    [EXAMPLE-5: PDB, `05_01_rt.xyz`, out of center -> Solomon link],

    Results https://knotgenom.cent.uw.edu.pl/compute/495a3b65deab878,

    [EXAMPLE-6: XYZ, `05_01_rt.xyz`, direct method-> Other],

    results https://knotgenom.cent.uw.edu.pl/compute/973ec36a39d340e,

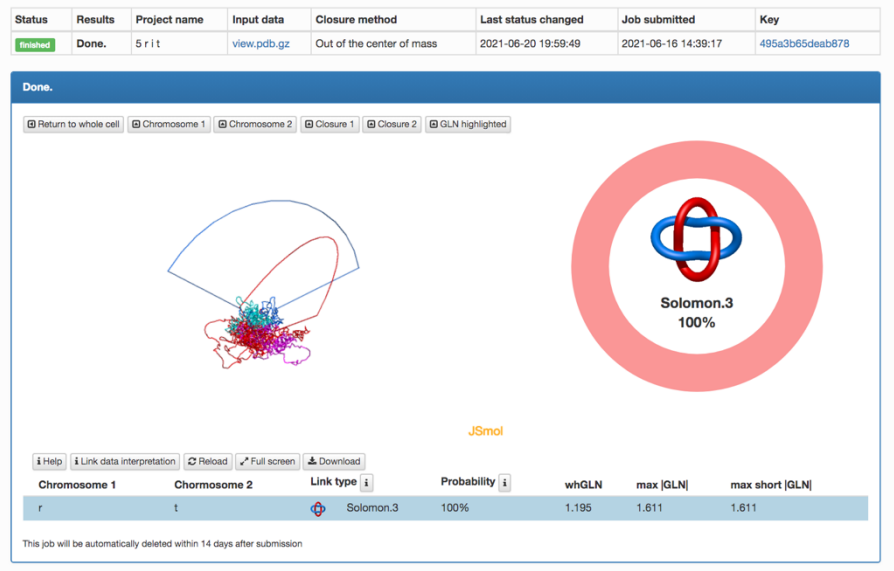
    [EXAMPLE-7: XYZ, `05_01_rt.xyz`, random method -> Other, Hopf style],

    Results https://knotgenom.cent.uw.edu.pl/compute/d3f7ca9f9f10e0a.

For whole cell, never mind if it is set of two or forty chains, the results are presented on such matrices:

On each output, Left: Interactive linking matrix: Diagonal elements describe knot types of individual chromosomes while lower off-diagonal elements characterize linking between chains. The respective color codes are displayed below the matrix. Middle and right: interactive matrices, which display whole GLNs and max|GLN|s.
Two GLN matrices are the same in all three outputs, since GLN is computed for open chains and does not depend on particular closure. However, knot and link types differ.

12. Below there is a list of untrivial knots and links that were found in the structure. To check the details for specific chain or pair of chains you can click “View details” butthon on the list, or just click on an appropriate cell on the matrix. Lets look at the link in “Out of the center of mass” method:

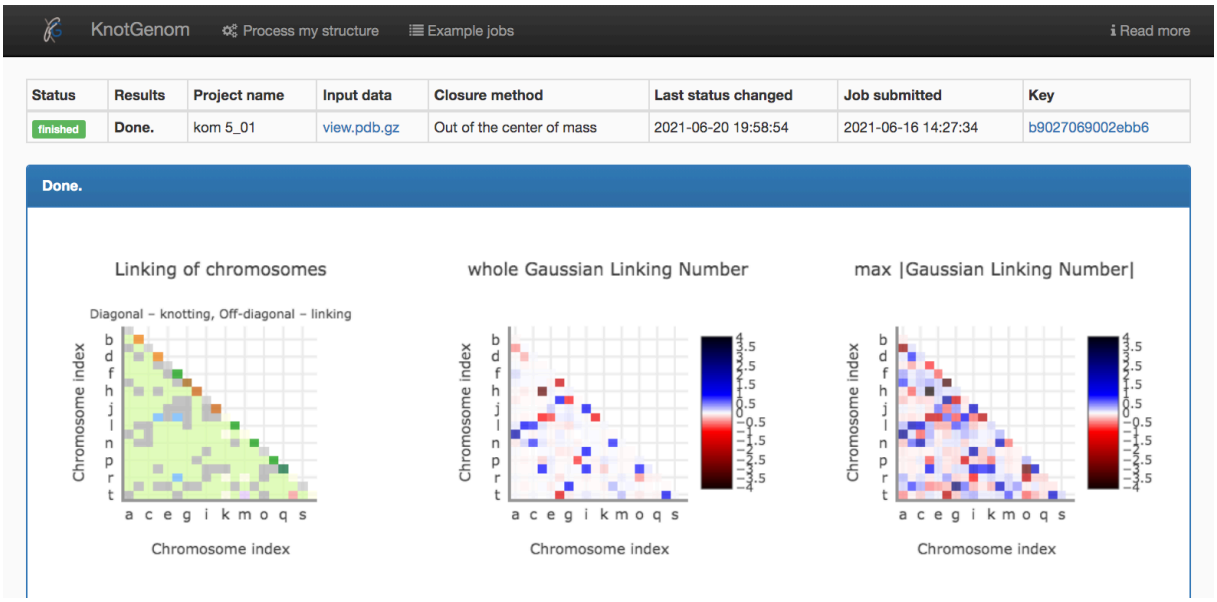


(Added arcs are linked, that’s why we have Solomon link, when whGLN is around 1. Try to click buttons over the structure to hide parts of the picture).

13. Whole cell (input file format: whole cell, XYZ):

    [EXAMPLE-8: XYZ, `05_01.xyz`, out of center],

    Results https://knotgenom.cent.uw.edu.pl/compute/b9027069002ebb6



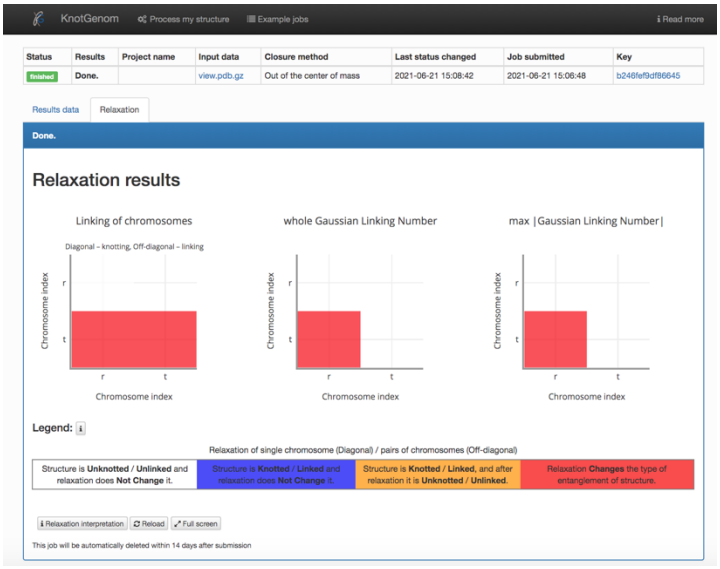
The same as for a pair, but more data and colors ☺ And small preview at the beginning.

Relaxation

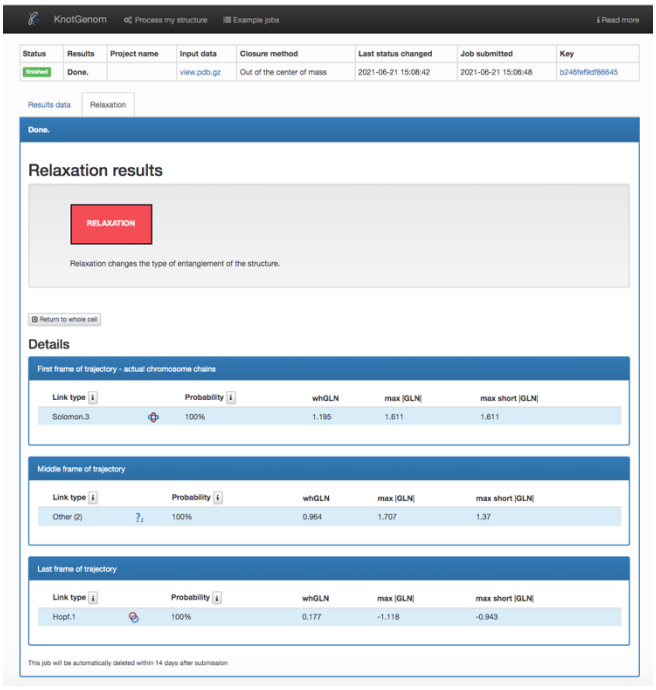
14. Pair of chromosomes r and t (input file format: whole cell, XYZ; relaxation: Yes):

- **[EXAMPLE-9: XYZ, `05_01_rt.xyz`, out of center, relaxation]**, Results https://knotgenom.cent.uw.edu.pl/compute/b246fef9df86645#resultstab

Click on “Relaxation” at the top.



Colors indicate candidates for stable knots or links. Blue depicts the stable complex topology. In this case link is unstable – link type and GLN values change along short relaxation - 10 frames are analyzed, clicking on the square on the matrix we can have a look at changing topology in three choosen frames:



15. Whole cell 5 with relaxation:

- **[EXAMPLE-10: XYZ, `05_01.xyz`, out of center, relaxation]**,

Results https://knotgenom.cent.uw.edu.pl/compute/6a3b88926e04551#resultstab

Examples:
Pair b-l: it unlinks, however link type method would suggest stability (Other…)
Pair i-j: maybe stable! (Hopf)

ENJOY KNOTGENOME by YOURSELF ☺