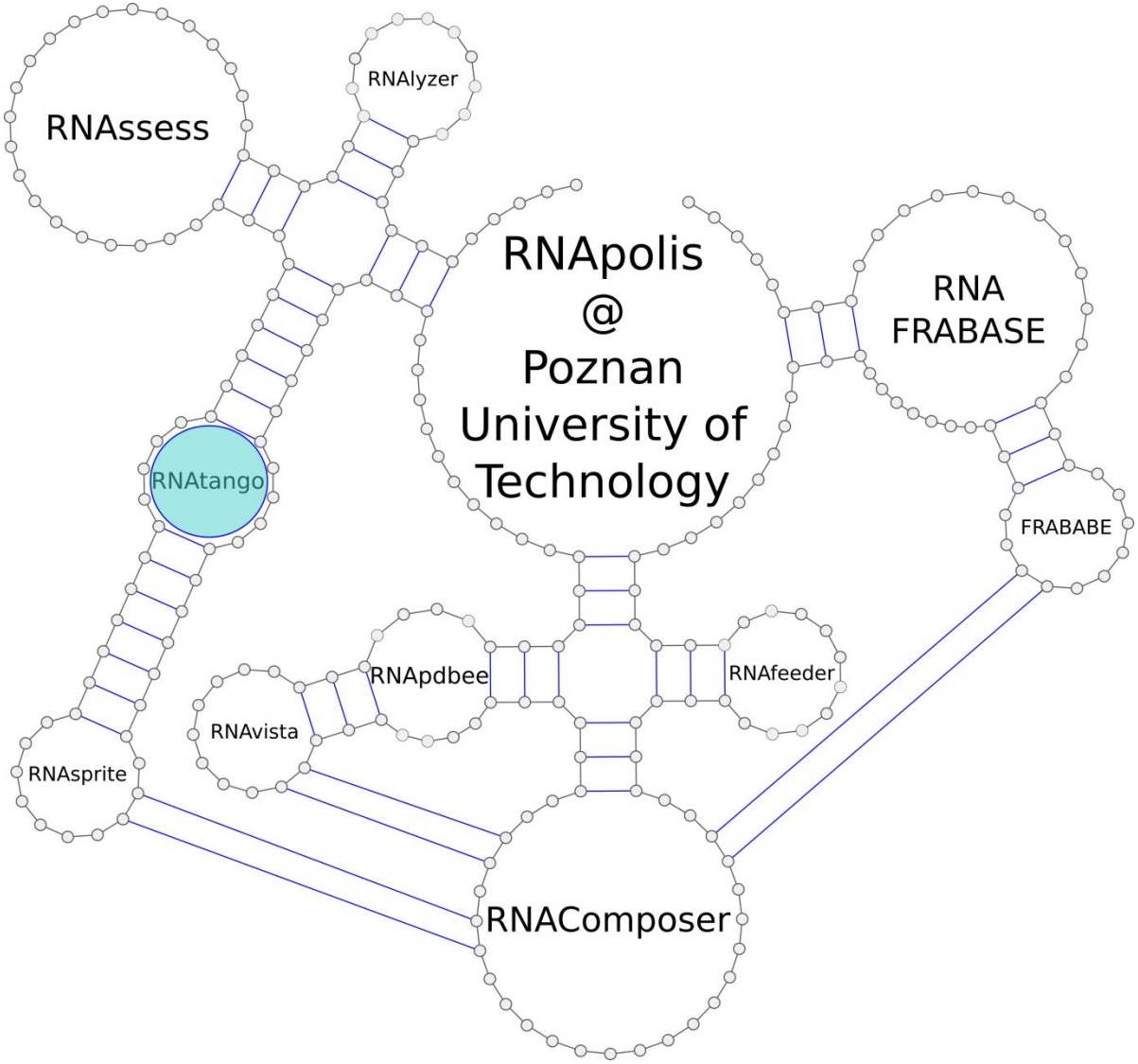
Torsion angle-based RNA 3D structure assessment in RNApolis

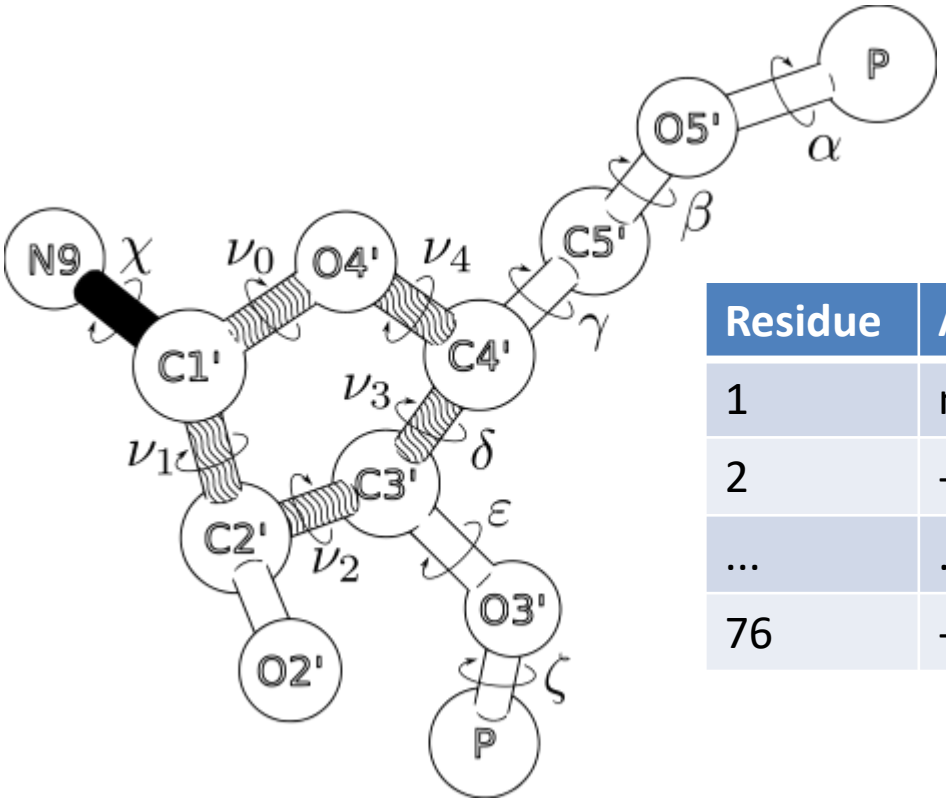
Tomasz Zok

Poznan University of Technology
& Institute of Bioorganic Chemistry, Polish Academy of Sciences

Torsion angle-based methods in RNApolis



Torsion angle space

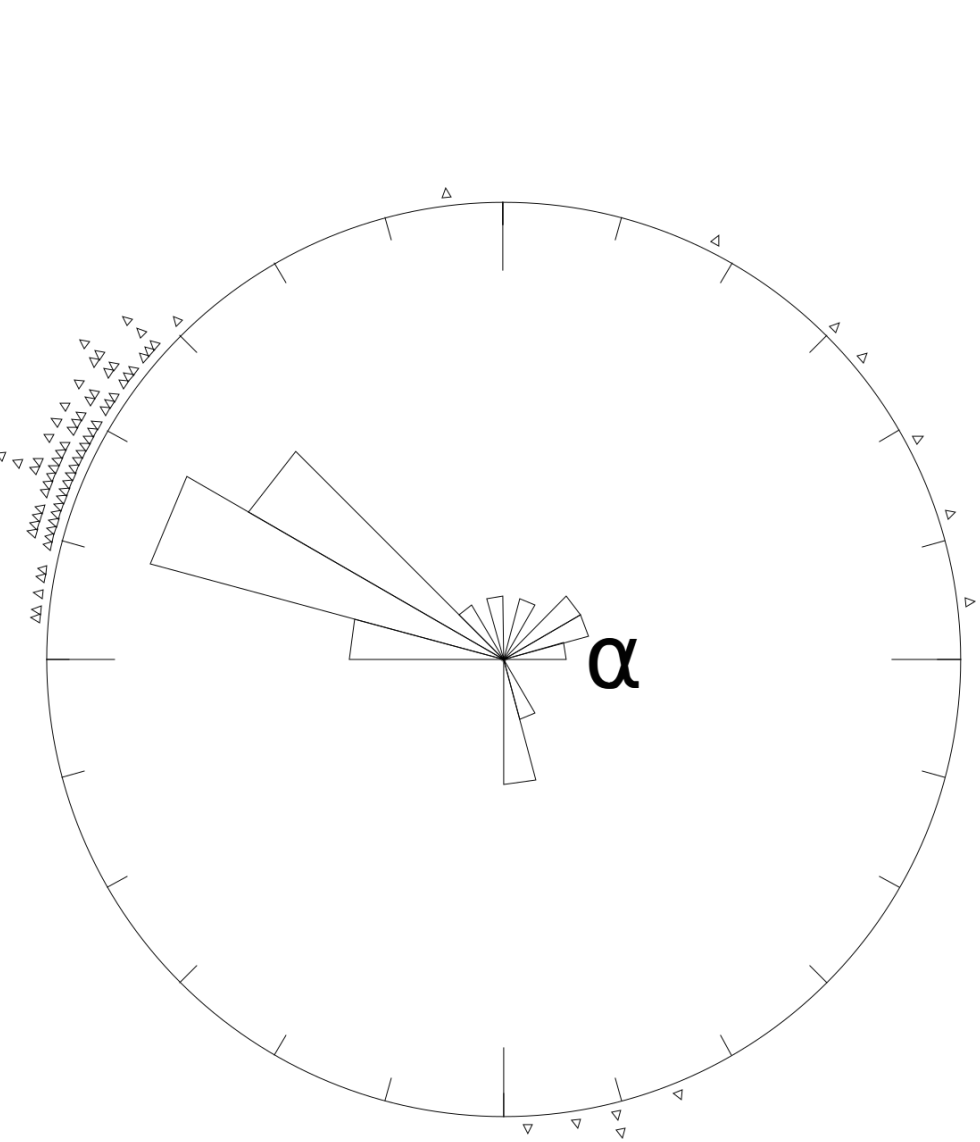


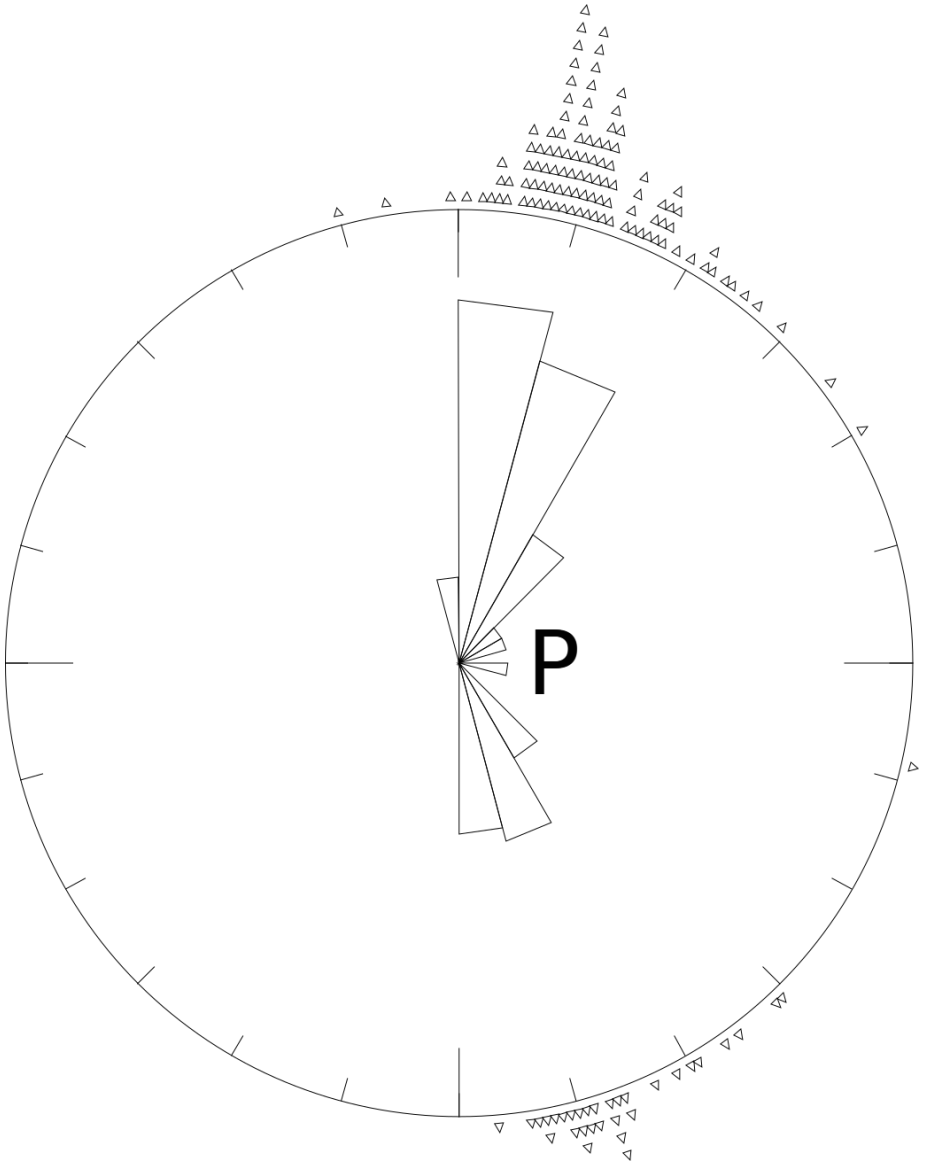
Residue	Alpha	Beta	...	Zeta
1	n/a	-128.1	...	-68.6
2	-67.4	-178.4	...	-76.8
...	...	...	...	...
76	-71.0	130.2	...	n/a

Motivation

- Values distribution in a sample for a given angle type:
  - Modality?
  - Mean, median?
  - Variance, skewness?
- Any missing values? Why?
- Any values outside of valid ranges? Why?
- How does torsion angles in one RNA correspond to those in another one? Similar? Dissimilar? What does it imply?

Angular histograms





Mean of Circular Quantities (MCQ)

$$\mathrm{MCQ}(\mathbf{S},\mathbf{S}') = \arctan\left(\frac{\mathbf{1}}{|\mathcal{R}||\mathcal{T}|}\sum_{\angle\in\mathcal{T}}\sum_{(\mathbf{r},\mathbf{r}')\in\mathcal{R}}\sin\boldsymbol{\Delta}(\angle_{\mathbf{r}},\angle_{\mathbf{r}'}),\frac{\mathbf{1}}{|\mathcal{R}||\mathcal{T}|}\sum_{\angle\in\mathcal{T}}\sum_{(\mathbf{r},\mathbf{r}')\in\mathcal{R}}\cos\boldsymbol{\Delta}(\angle_{\mathbf{r}},\angle_{\mathbf{r}'})\right)$$

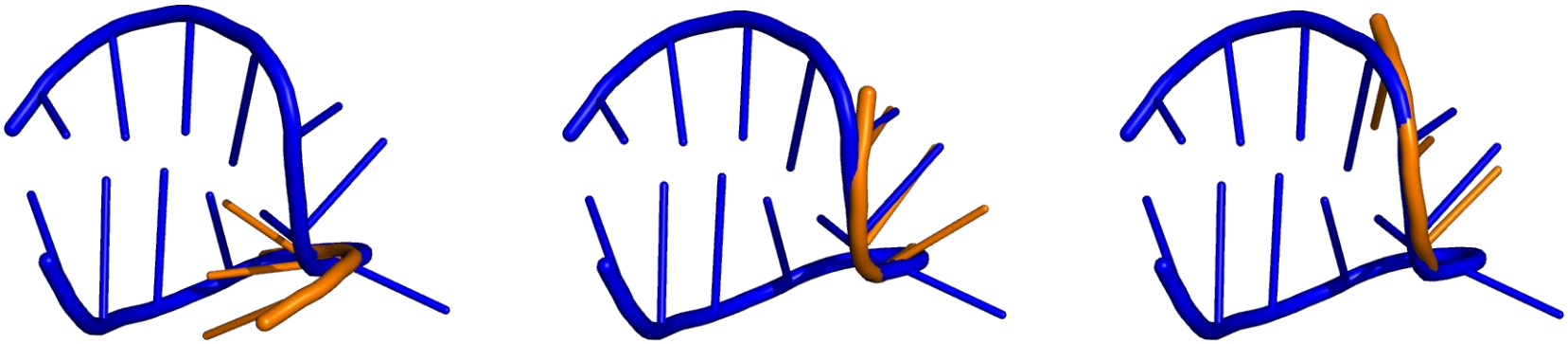
Where:

- $\mathbf{S}$, $\mathbf{S}'$ – two compared structures
- $\mathcal{R}$ – a set of residue pairs $\{(r, r') \in S \times S'\}$
- $\mathcal{T}$ – a set of torsion angles
- $\angle_r$ – a value of torsion angle $\angle$ in residue r
- $\boldsymbol{\Delta}(\angle_{\mathbf{r}}, \angle_{\mathbf{r}'})$ – subtraction operation handling the periodicity of angles

Structural alignment

- Structures S and S' can be aligned using MCQ with a threading-like optimization process
- Let's assume that *|S| = length(S)* and *|S'|=length(S')* and *|S|>|S'|*
- Let's denote as *score(**S, S', n**)* an MCQ value between residues in **S** and **S'** starting from **n**-th nucleotide of **S**
- An optimal alignment w.r.t. MCQ distance measure is described by a value of **n***:

$$n^* = \underset{0<i<|S|-|S'|}{\arg\min} \; score(S, S', i)$$



Chain matching

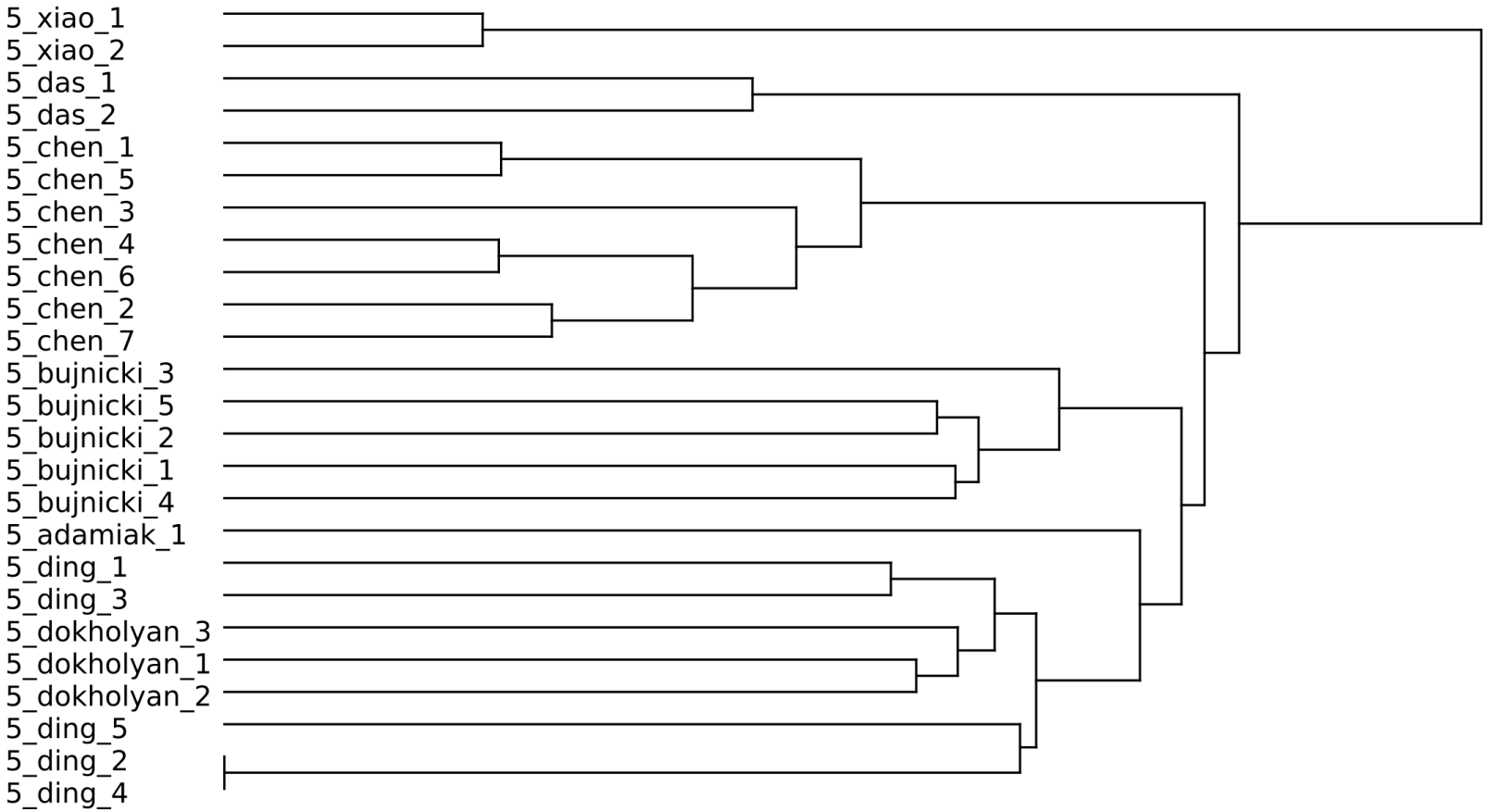
- Often a single PDB file contains many chains with varying names
- Without *a priori* knowledge about them, how can one find a way to match chains from structures **S** and **S'**?
- Let's denote $\mathbf{n^*_{u,v}}$ as a value of **n*** (optimal alignment) between chain **u** of structure **S** and chain **v** of structure **S'**
- A decision variable $\mathbf{x_{u,v}}$ is set to **1** if chain **u** and **v** are selected to be matched, and **0** otherwise
- An optimal matching is one that minimizes the formula:

$$\sum_{u \in S} \sum_{v \in S'} n^*_{u,v} x_{u,v}$$

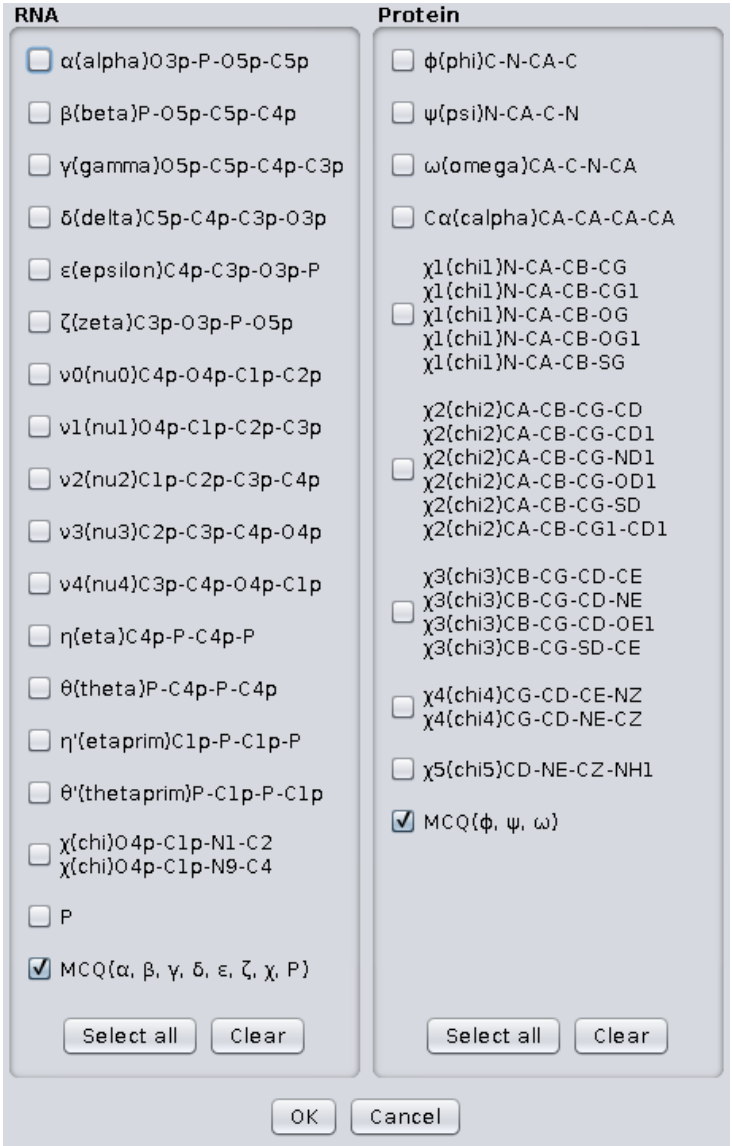
- This is a typical case of **Assignment Problem**

Global comparison

- MCQ used as a distance measure can produce a **dissimilarity matrix**
- The data can be visualized or analyzed further

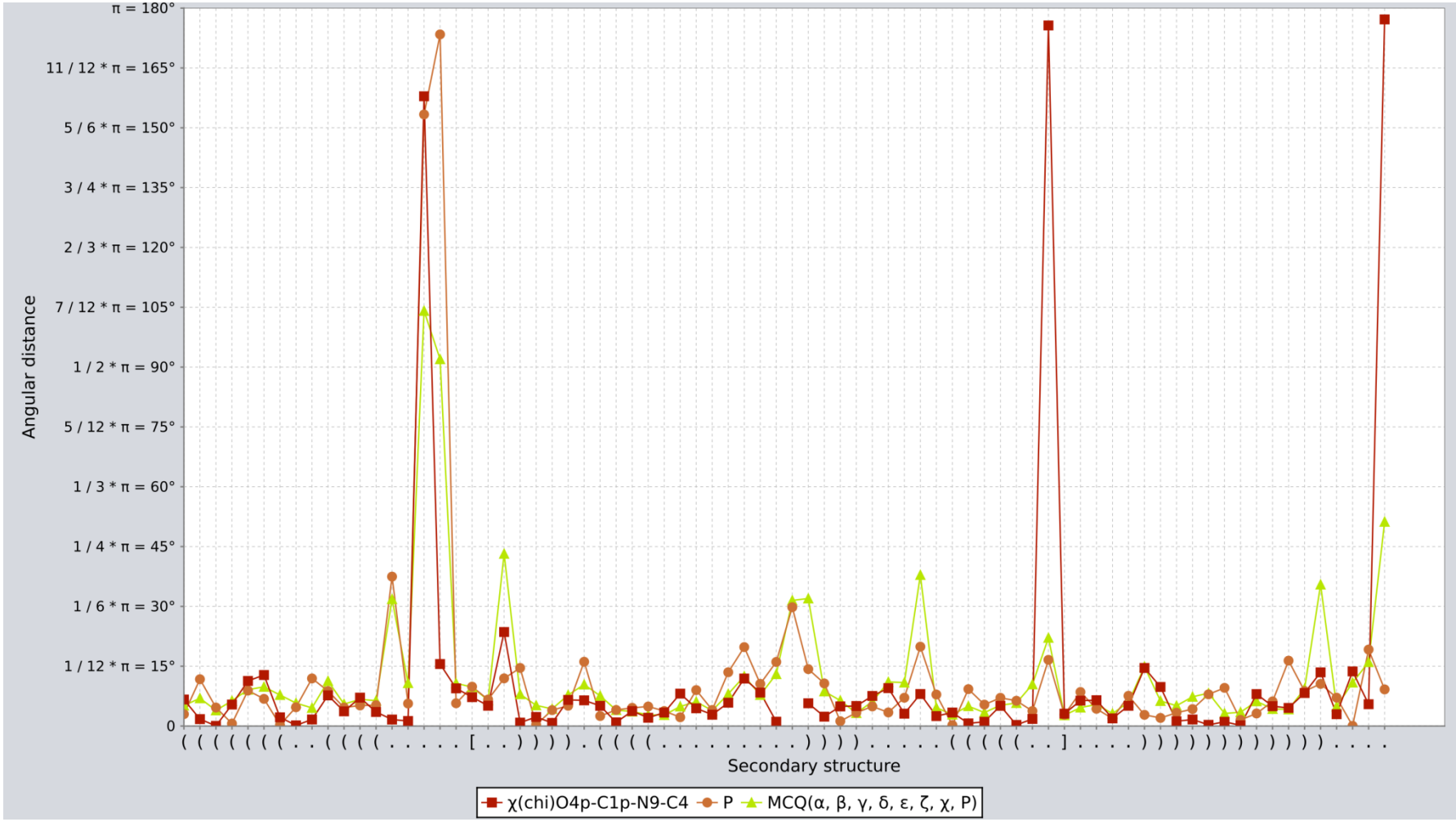


Local comparison

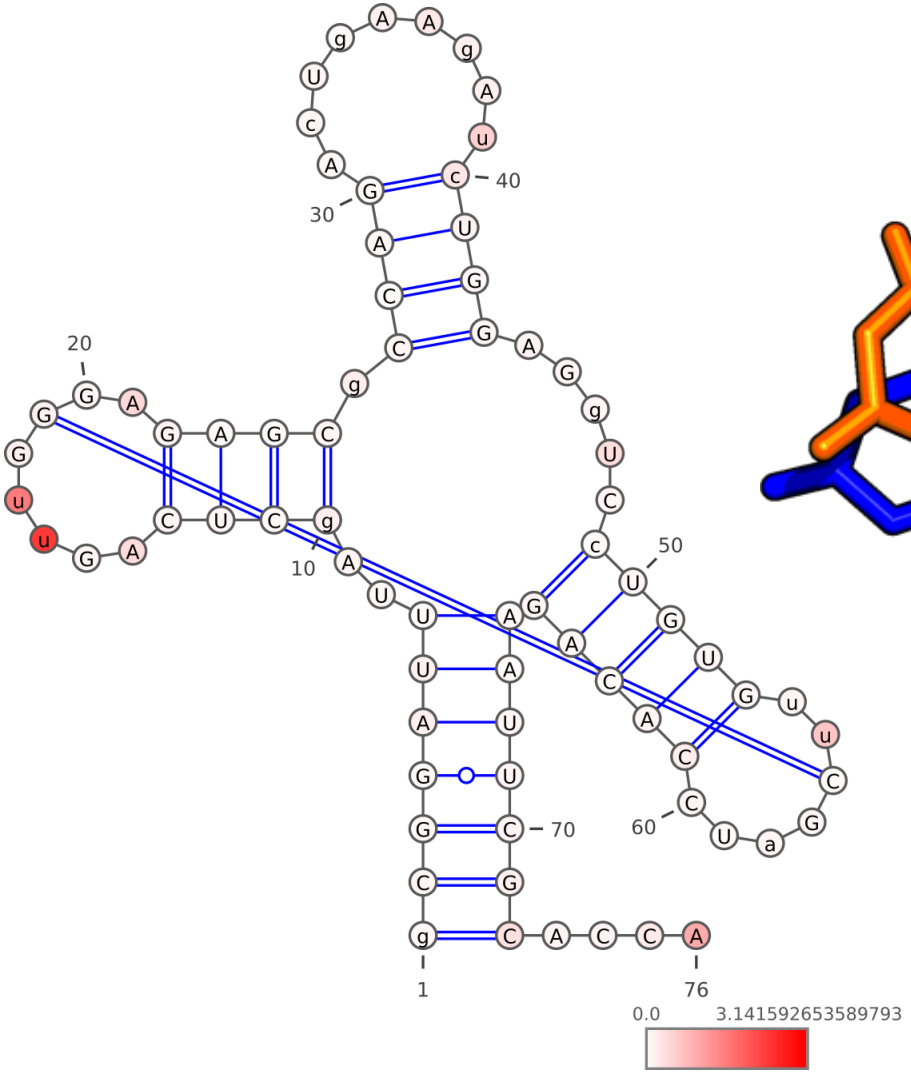


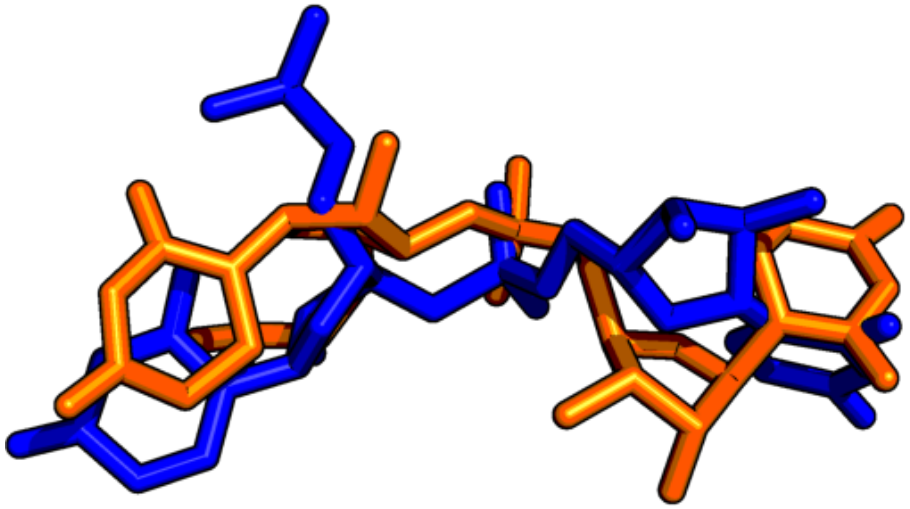
- For a pair of structures, a more in-depth and better parameterised results may be obtained
- Distances may be calculated on specific angles only
- The chain matching and structural alignment may be computed on a subset of angle types

Local comparison

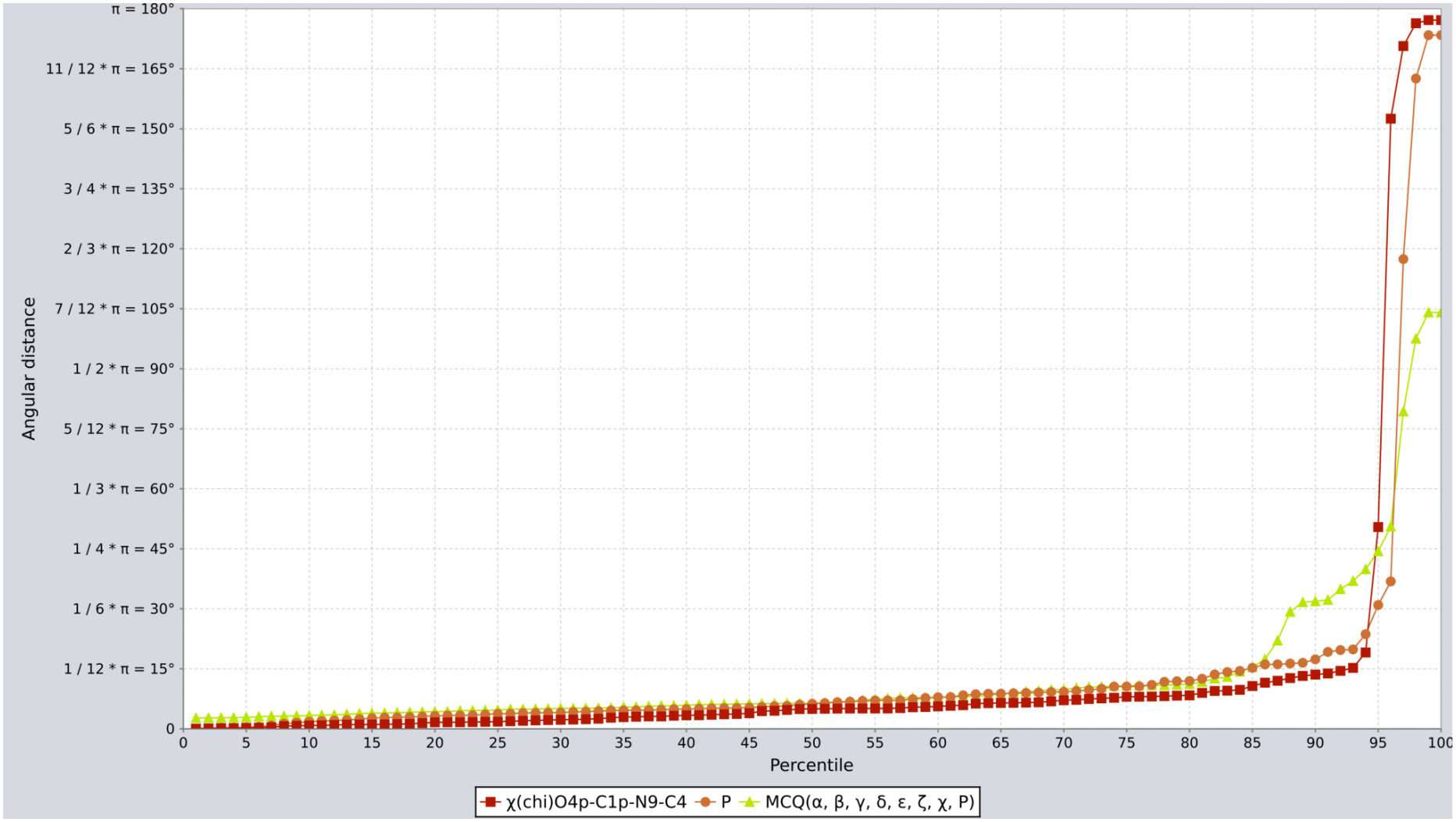


Local comparison

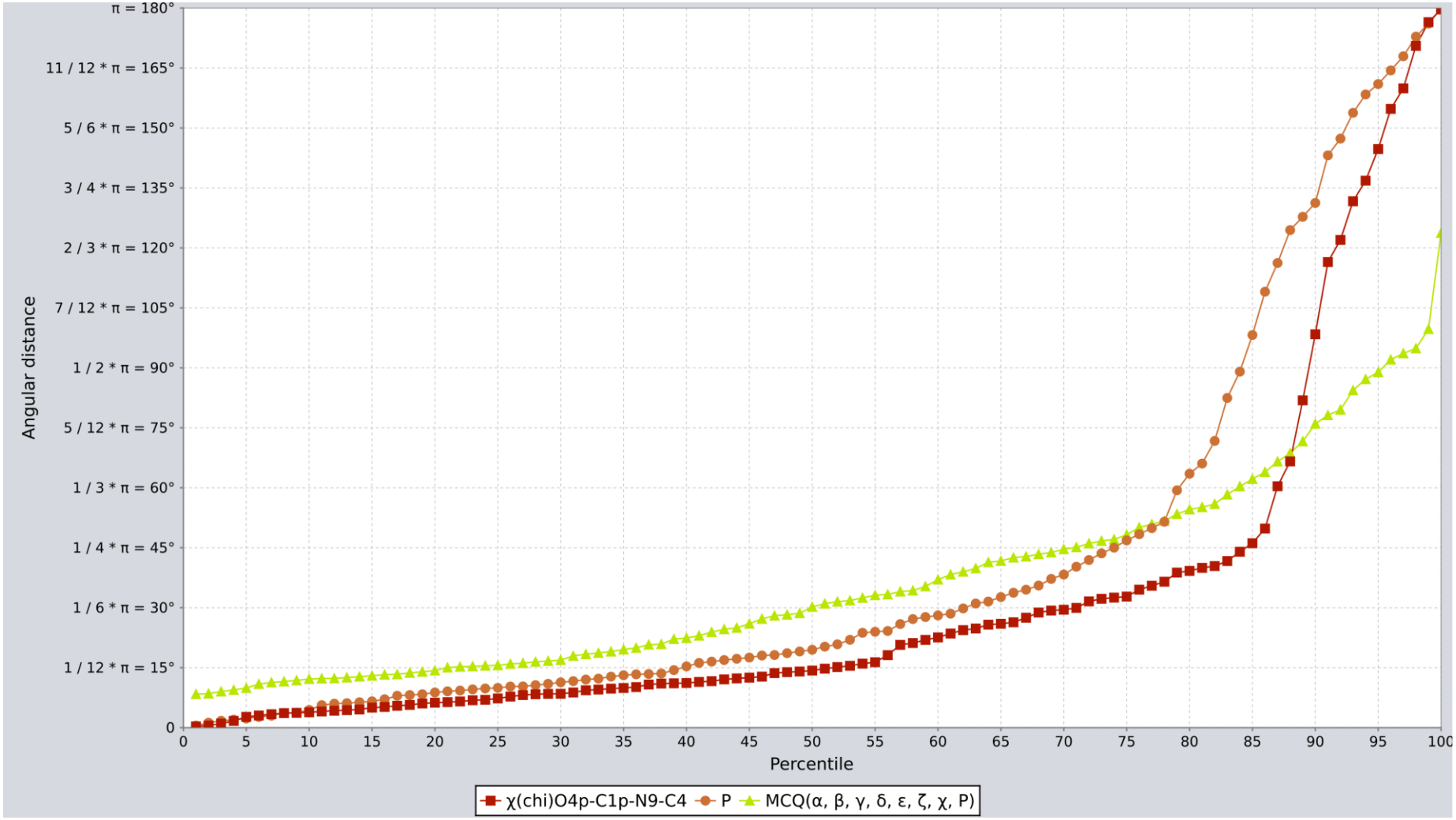




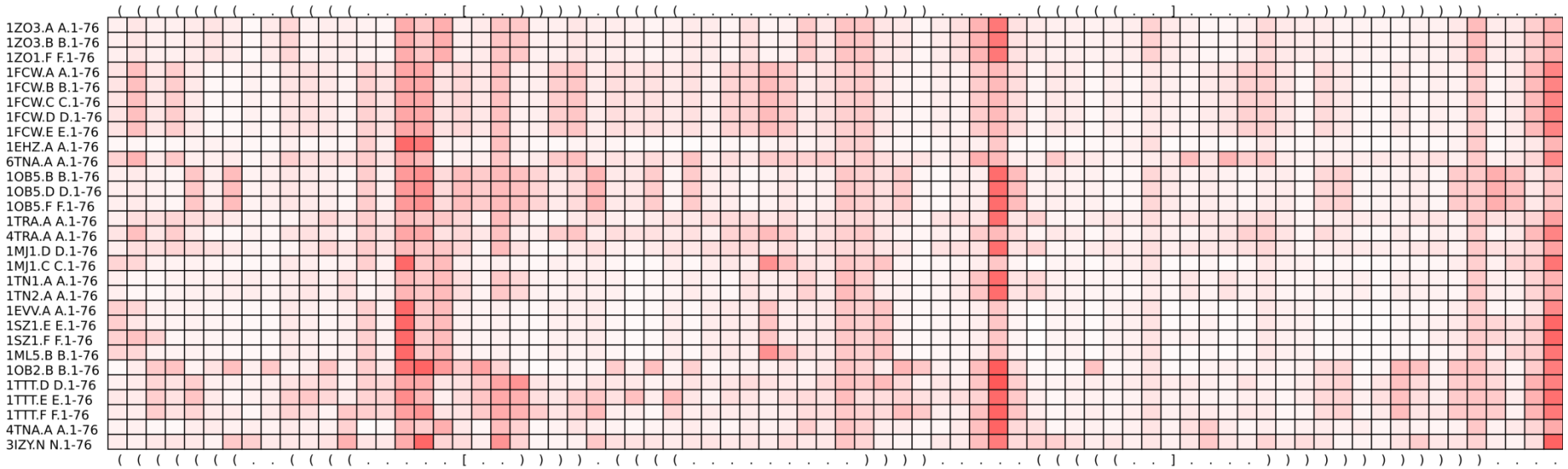
Local comparison



Local comparison

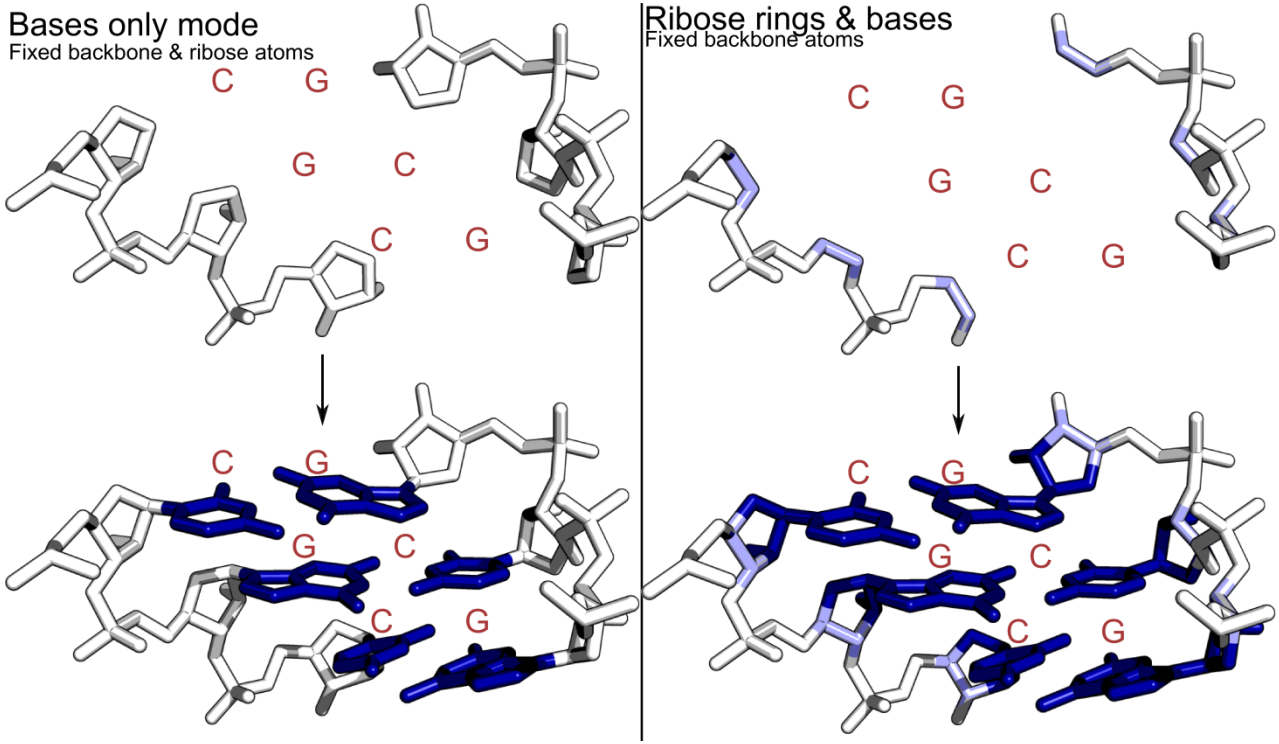


Local comparison



Nucleotide library

- **Input:** a set of vectors of torsion angles, each for a single nucleotide
- **Output:** a collection of representative nucleotides
- **Practical motivation:** the library can be used to redesign RNA chain by replacing its nucleotides' fragments with those from representative conformer



Future ideas

- Given sequence and secondary structure, predict the values of torsion angles
- For a set of reference, high-quality structures compute a statistical knowledgebase such as distributions, or correlations; use it to score other RNA structures and find their shortcomings
- Classify spatial contacts in terms of the torsion angles between base/ribose/phosphate

Conclusions

- Torsion angles can be used to create an alternative representation of RNA 3D structure
- This allows to propose novel ways to analyze RNA structure
- Circular statistics of torsion angles alone or their differences is an indispensable tool
- It forms a basis for structure comparison in torsion angle space and its many applications
- One of them being the construction of nucleotide library useful among others in RNA redesign
- Even more ideas come to mind and new approaches await to be checked

Software for the analysis of RNA structures in torsion angle space is available at:

http://www.cs.put.poznan.pl/tzok/mcq/