Comparison of multiple structures in torsion angle space

Tomasz Zok[1], Mariusz Popenda[2], Marta Szachniuk[1,2]
Contact: Tomasz.Zok@cs.put.poznan.pl

[1] Institute of Computing Science, Poznan University of Technology
[2] Institute of Bioorganic Chemistry, Polish Academy of Sciences

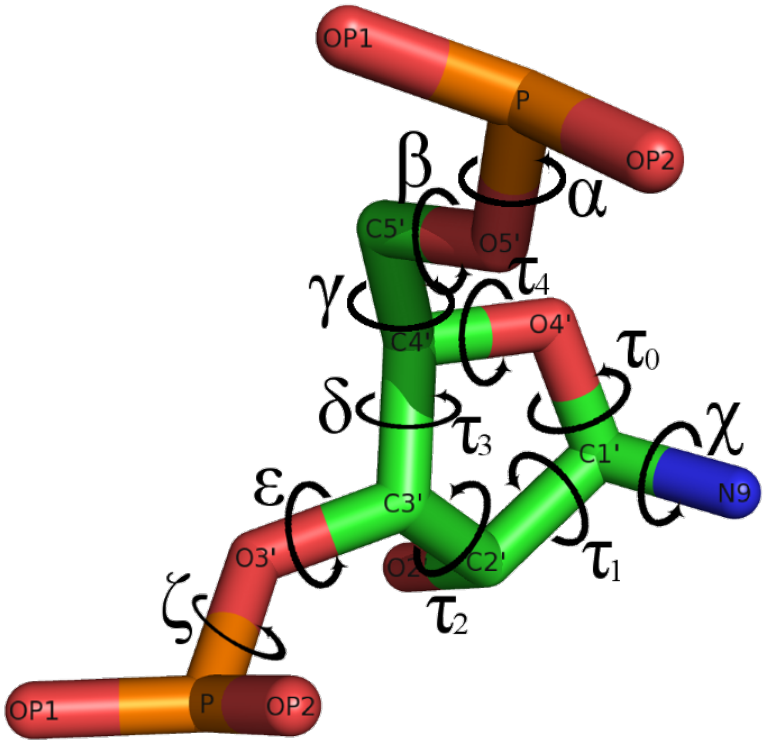
28.03.2013

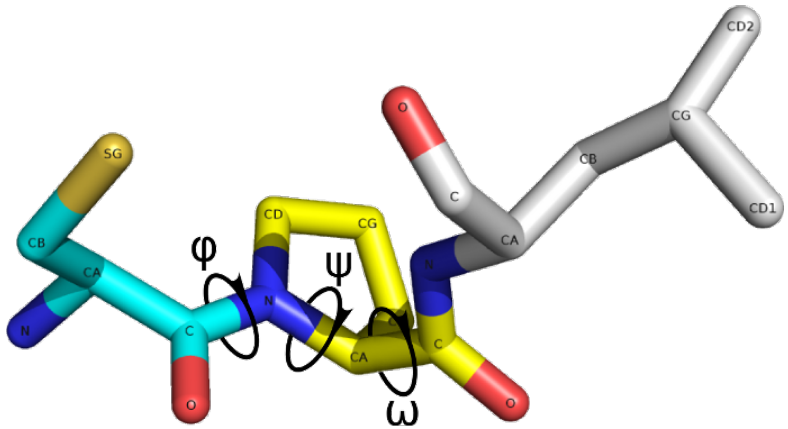
Introduction

Comparison of biological molecule structures (e.g. RNAs, proteins) provides an insight into the details of their folds. A crucial element of the comparison process is the choice of similarity metrics. Structure representation in torsion angle space proves to be very useful for this purpose [1], with MCQ as a distance measure. Here, we present how it can be incorporated into the scenario of multiple structure comparison and we show the visualization of MCQ application for two use cases.

Torsion angle space

The 3D shape of a single residue (left figure: nucleotide, right figure: amino acid) can be described by a set of torsion angles, which completely define its structure.





Torsion angle space

The overall molecule structure can be thus represented by a set of torsion angle values provided for each residue (cf. table with selected information about 1EHZ – 76-nucleotide t-RNA).

resid	α	β	γ	δ	ε	ζ	χ	P
1	-	-128.1	67.8	82.9	-155.6	-68.6	-167.8	16.1
2	-67.4	-178.4	53.8	83.4	-145.1	-76.8	-163.8	16.1
...								
76	-71.0	130.2	164.6	160.9	-	-	138.5	176.1

Mean of Circular Quantities (MCQ)

Mean of circular quantities is an average of circular (e.g. angle) values. In our approach it can be used as a distance measure to compare molecule structures in torsion angle space. The following formula is applied:

$$\mathrm{MCQ}(\mathbf{S},\mathbf{S}') = \arctan\left(\frac{1}{\mathbf{r}|\mathbf{T}|}\sum_{\mathbf{a}\in\mathbf{T}}\sum_{\mathbf{r}}^{\mathbf{i=1}}\sin\boldsymbol{\Delta}(\mathbf{a_i},\mathbf{a_i'}), \frac{1}{\mathbf{r}|\mathbf{T}|}\sum_{\mathbf{a}\in\mathbf{T}}\sum_{\mathbf{r}}^{\mathbf{i=1}}\cos\boldsymbol{\Delta}(\mathbf{a_i},\mathbf{a_i'})\right)$$

where:

- $\mathbf{S}$, $\mathbf{S}'$ – two compared structures
- $\mathbf{r}$ – number of residues (the same for $\mathbf{S}$ and $\mathbf{S}'$)
- $\mathbf{T}$ – a set of torsion angles
- $\boldsymbol{\Delta}(\mathbf{a_i},\mathbf{a_i'})$ – function which takes into account the periodicity of angles

MCQ parametrization

Case 1: **r** = all residues, **T** = all predefined angles

Result = a single MCQ value for a pair of structures

A global dissimilarity score

Allows to build a dissimilarity matrix in a many-to-many comparison

Case 2: **r** = 1, **T** = all predefined angles

Result = a vector of MCQ values (1 value for a pair of residues)

Allows to find fragments of high or low similarity

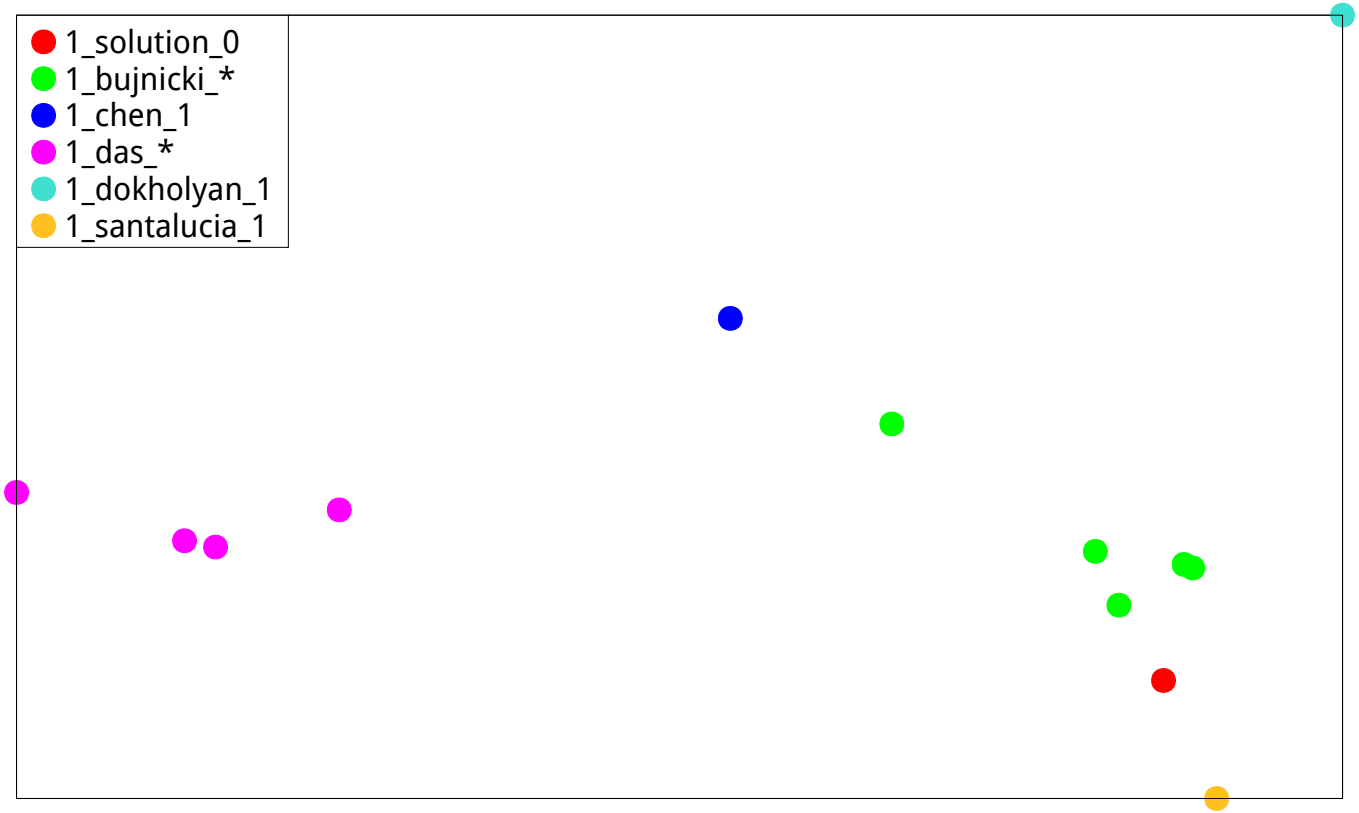
Use case I: models from RNA-Puzzles

As the first use case, we have chosen RNA-Puzzles challenge 1 (dimer):

> Predict the structure of the following sequence: 5'-CCGCCGCGCCAUGCCUGUGGCGG-3' knowing that the crystal structure shows a homodimer that contains two strands of the sequence that hybridize with blunt ends (C-G closing base pairs) [2].

Use case I: models from RNA-Puzzles

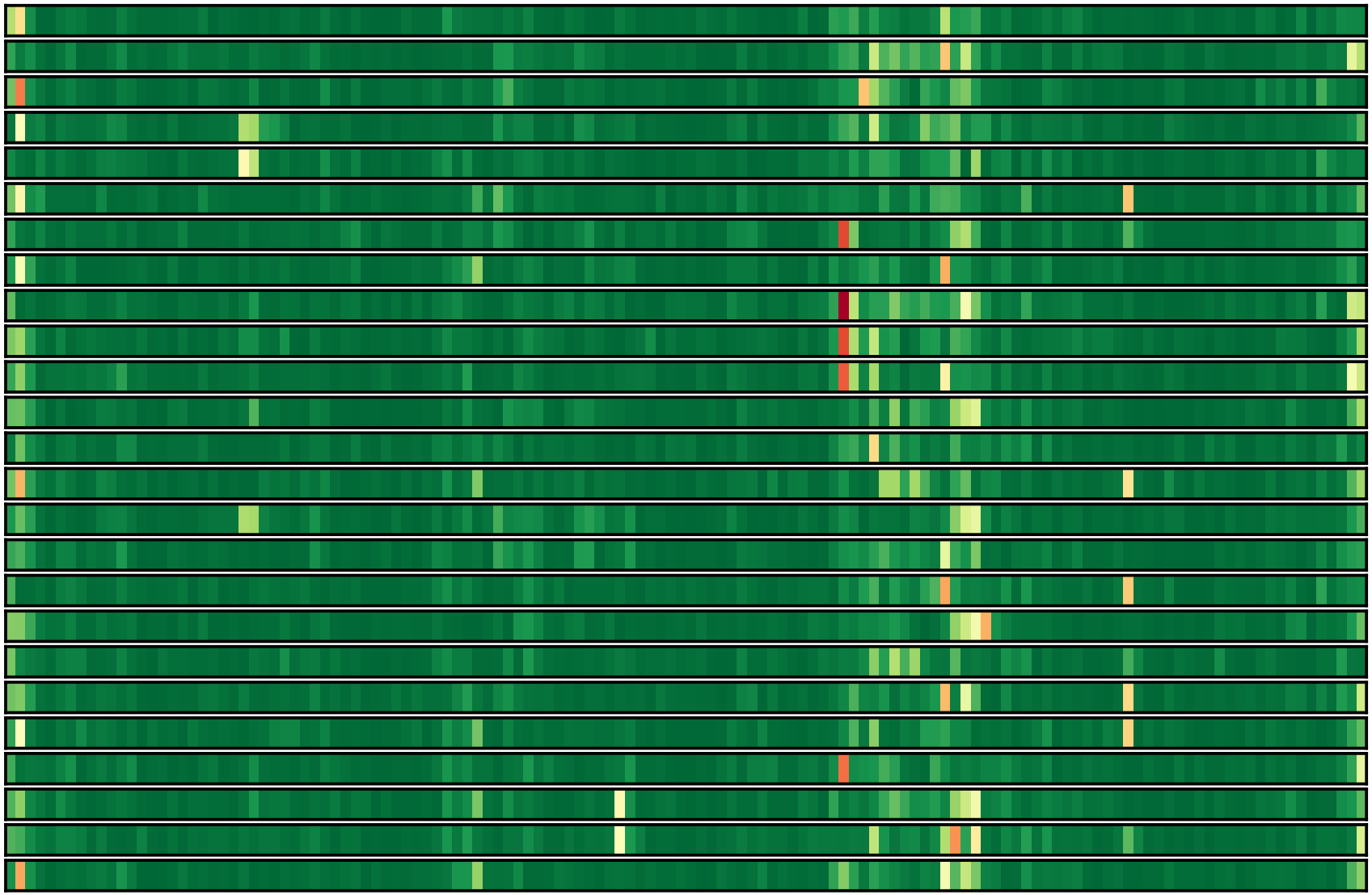
A dissimilarity matrix of all submitted models was calculated and transformed into 2D visualization (classical multidimensional scaling) to show relative distances in torsion angle space.



Use case II: NMR models

The second use case concerns 1M2F associated with 25 NMR 3D models [3]. Their comparison result is visualized as 2D colormap, where the vector of MCQ values is colored according to the distance between residues:

Use case II: NMR models



Software availability

The executable and source code licensed under BSD are freely available and can be downloaded from:
`http://www.cs.put.poznan.pl/tzok/mcq/`

References

T. Zok, M. Popenda, and M. Szachniuk.

MCQ4Structures to compute similarity of molecule structures.

CEJOR, 2013, DOI: 10.1007/s10100-013-0296-5

J. A. Cruz et al.

RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction.

RNA, 18:610–625, 2012.

S. B. Williams, I. Vakonakis, S. S. Golden, and A. C. LiWang.

Structure and function from the circadian clock protein KaiA of Synechococcus elongatus: a potential clock input mechanism.

Proceedings of the National Academy of Sciences of the United States of America, 99(24):15357–62, 2002.