

## Turns revisited: Clustering turn structures using ESOMs leads to a uniform classification for open, normal and reverse turn families.

O. Koch<sup>1/2</sup>, G.. Klebe<sup>2</sup>

<sup>1</sup> *The Cambridge Crystallographic Data Centre, Cambridge, CB2 1EZ, UK*

<sup>2</sup> *Philipps-University Marburg, Institute of Pharmaceutical Chemistry, 35032 Marburg, Germany  
Email: Mr.O.Koch@Gmail.com*

In contrast to helices and  $\beta$ -sheets, turns are irregular secondary structure elements. They are up to six residues in length and contain a hydrogen bond or a specific C $\alpha$ -C $\alpha$  distance between the first and last residue. Because of their irregularity and a lack of data in the past, previous classifications do not accommodate possible new turn-types in the current protein structures. Additionally, there is a lack of an overall classification for all turn families. Therefore, a new classification was done from scratch. Based on a non-redundant dataset of 1903 protein chains, all possible turn structures were retrieved from Relibase<sup>[1]</sup> using Reliscript (the Python-based interface) and clustered using Emergent SOMs<sup>[2]</sup>. The backbone torsion angles describing a turn, including the  $\omega$  torsion angle, were used as the feature vector.

In general, a hydrogen bond between CO<sub>i</sub> – NH<sub>i+n</sub> is expected<sup>[3]</sup> within hydrogen bonded turns ('normal' turns, Figure 1b), but this analysis shows that NH<sub>i</sub> – CO<sub>i+n</sub> hydrogen bonded turns also exists ('reverse' turns, Figure 1a). They have been theoretically described<sup>[4]</sup>, but never reported previously in proteins. Furthermore, a C $\alpha$ -C $\alpha$  distance cut-off of 10 Å was chosen for structures lacking a hydrogen bond followed by a visual inspection of the retrieved clusters to identify turn structures ('open' turns, Figure 1c).

An analysis of these turn families reveals that, based on the amino acid propensities, the differentiation into normal, open and reverse turn families seems reasonable. Additionally, a large fraction of open turn-types would be ignored using a shorter distance cut-off. Finally, this survey describes 3 *open* turn, 4 *normal* and 5 *reverse* turn families with several turn-types that have not been previously described (Table 1). Protein sequence-based turn prediction with high accuracy confirmed this new categorization based on machine learning methods as consistent and well-defined<sup>[5]</sup>.

In addition to the information about helices and  $\beta$ -sheets retrieved from the PDB, this new uniform classification of turn families is integrated into Secbase, a new extension of Relibase. Relibase is an object-oriented data management system and stores the three dimensional structural information of protein-ligand complexes deposited in the PDB. Both tools provide integrated access for the analysis of secondary structure elements within proteins, protein-protein interfaces and ligand binding.

1. Hendlich, M.; Bergner, A.; Günther, J.; Klebe, G. Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.* **2003**, 326(2), 607-620.
2. Ultsch, A. Maps for the Visualization of high-dimensional Data Spaces. In *Proceedings of Workshop on Self-Organizing Maps (WSOM)*; Kyushu, Japan, **2003**, 225-230.
3. Chou, K.C. Prediction of tight turns and their types in proteins. *Anal. Biochem.* **2000**, 286, 1-16.
4. Toniolo, C. Intramolecularly hydrogen-bonded peptide conformations. *CRC Crit. Rev. Biochem.* **1980**, 9(1), 1-44.
5. Meissner, M.; Koch, O.; Klebe, G.; Schneider, G. Prediction of Turns in Protein Structure by Kernel-Based Machine Learning Methods, *submitted*

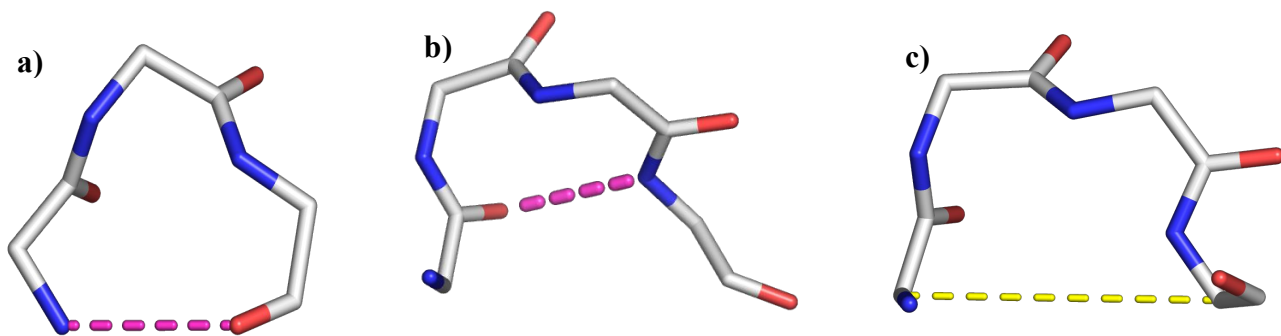


Figure 1: Different turn conformations: a) reverse  $\epsilon$ -turn (3 residues), b) normal  $\beta$ -turn (4 residues), c) open  $\beta$ -turn (4 residues)

Table 1: Description of used datasets for clustering with number of retrieved turn-types (open turns show structure cluster that looks more like: a) a kink or b) a hook)

	designatio n	number of		
		residues	structures	turn-types
open	$\beta$	4	137101	11 + 6 <sup>a</sup>
	$\alpha$	5	19607	21 + 1 <sup>a</sup>
	$\pi$	6	21204	22 + 6 <sup>b</sup>
normal	$\gamma$	3	20198	2
	$\beta$	4	28718	6
	$\alpha$	5	91726	9
	$\pi$	6	3994	8
reverse	$\delta$	2	210	9
	$\epsilon$	3	134	5
	-	4	1957	18
	-	5	1340	21
	-	6	954	13