

# Discussion on Occurrence Regularity of Cis - peptide

Yali Zhang<sup>1</sup>, Yuyun Wu<sup>1</sup>, Hongying Yang<sup>1</sup>, Shuhui Yu<sup>2</sup>, Li Zhao<sup>1</sup>, Yingbin Yang<sup>1</sup>

<sup>1</sup>Southwest University, Chongqing, 400715

<sup>2</sup>Chongqing University, Chongqing, 400044

**Keywords:** Cis-Peptide Bond; Incidence; B-Catenin; Nuclear Accumulation

**Abstract.** In order to investigate the occurrence of cis-peptide bond, the study counted the incidence of different cis-peptide bonds from a non-redundant data set containing 1591 protein PDBs. Then, 20 different amino acid residues were used to replace  $\beta$ -catenin S246 residues and transferred into HepG2 cells and observed the variation of the storage of different mutants in the nucleus and the interaction with APC. The results showed that the cis-peptide bond of 82.30% (703/854) was distributed in XP X is an arbitrary amino acid residue, P is a proline residue) site of the protein polypeptide chain, and different cis-peptide The incidence of different, which when X is tyrosine, tryptophan and other residues XP has the highest incidence of cis-peptide Bond; and X is leucine, methionine and other residues XP has the lowest results of the interaction between APC and  $\beta$ -catenin was weak when S246 mutated into tyrosine and tryptophan residues, and the interaction of APC with  $\beta$ -catenin was weak. When S246 mutates into leucine and methionine residues, the interaction between APC and  $\beta$ -catenin is relatively enhanced. It is suggested that the change of  $\beta$ -catenin function is related to the change of cis-peptide bond In the X246P site. The results of this study are expected to provide new strategies for genetic disease prediction, genetic diagnosis, and protein modification.

## Introduction

The N-C $\alpha$  bond, C $\alpha$ -C $\beta$  bond, and C = N (peptide bond) form the backbone of the protein polypeptide chain, which is composed of 20 amino acid residues of the human protein. The N-C $\alpha$  bond and the C $\alpha$ -C $\beta$  bond are a single bond in a peptide plane of the primary structure, so they are free to rotate the dihedral angles  $\Psi$  and  $\phi$  of the rotating axis. At the same time, since the peptide bond has a property of a partial double bond, the dihedral angle  $\omega$  with the peptide bond as the rotational axis can only be restricted to 0° or 180° [1]. 1970 The International Union of Theoretical and Applied Chemistry (IUPAC) defines a transpeptide bond when the dihedral angle is equal to 180° and is defined as a cis-peptide bond [2] when  $\omega$  is equal to 0°. Correspondingly, the conformation dominated by transpeptide bonds is called trans-conformation, and the conformation dominated by cis-peptide bonds is called cis-conformation. The cis-conformation and trans-conformation can be changed under certain conditions, called cis-trans isomerization. Human protein Amino acid polypeptide chain in more than 99% of the C = N bond is a trans peptide bond, and only less than 1% of the cis peptide bond [3]. Because the cis-isomerization of proteins is a slow allosteric, it plays an important role in the regulation of protein structure and function changes. Studies have shown that many life-activity regulation processes, such as regulation of protein stability [1, 3], Protein interactions [4,5], membrane protein binding [6,7], ion channels [8], cellular signaling pathways [9,10], gene expression [11], cancer [12], amyloid Formation [13,14] and Alzheimer's disease [15,16] and so on have an important relationship with the occurrence of protein cis conformation. Therefore, the characteristics of the protein cis peptide bond in the three-dimensional structure of the protein and the frequency (incidence) of the amino acid residues of the different cis-peptide bonds are studied. The structure of the amino acid residues that have been replaced by additional amino acid residues and the law of functional change has important significance.

## Materials and Methods

Cis-peptide bond definition and no redundancy PDB data set establishment: We will be cis-peptide bond is defined as:  $0^\circ < \omega < 90^\circ$ , the trans peptide bond is defined as  $90^\circ < \omega < 180^\circ$  range. According to the PDB file released by the RCSB protein Data Bank (PDB) on March 23, 2017, a human protein with a resolution of less than 0.20 nm and an identity of less than 30% was obtained using the X-ray experimental method PDB files (1591 in total) constitute the data set for this study.

X-Pro sequence characteristics affect the occurrence of cis-peptide bond analysis: In order to further describe the abundance of differences reflected by the cis-conformational tendency of the cis, we introduced the Z score, expressed by the formula:  $Z_i(a) = [X_i(a) - \mu] / \sigma$ , where  $X_i(a)$  is the number of residues a at the i-th position,  $\mu$  is the mean, and  $\sigma$  is the standard deviation. See the literature [11, 17] for specific methods.

Non-proline cis-peptide bond (X-XnP) statistics and analysis: first all cis-peptide bond (XX) to remove XP peptide bond and then statistics of different X-XnP cis-peptide bond type, number and Incidence, and get the relevant cis-occurrence of the law.

Carrier Construction and Lentiviral Packaging: To further investigate the relationship between the occurrence of cis-peptide bond and protein function in eukaryotic cells, we digest the BamHI and NotI double- $\beta$ -catenin fragments from plasmid human  $\beta$ -catenin pCDNA3.1 ( And then transferred to the lentiviral vector pCDHpCDH-CMV-EF1-copGFP-2A-Puro (purchased from: Chinese Academy of Sciences, Wuhan virus), and then use the direct point mutation method, respectively, cis-peptide bond incidence The highest and lowest two XP-substituted vectors encode the (S246P) site of the beta-catenin protein. Four different mutant vectors were then packaged into different lentiviral particles according to the method described in [18, 41], as shown in Fig.

Cell culture and stable expression lines: HepG2 (purchased from ATCC) culture method See [19], observed in the 12-well plate to 50-60% fusion, according to the literature [20] described the above method Packaging virus particles transfected cells. After 48 hours, the medium containing 10  $\mu$ g / mL of puromycin was added for resistance screening, and the medium was replaced once a week. The expression of the fusion protein copGFP was observed by inverted fluorescence microscopy. When cells were observed to grow to about 60% of the cells, most of the cells expressed copGFP, the cells were transferred to 100 mm dishes to expand the culture.

Cell Protein Extraction and Interaction Analysis: When HepG2 cells were grown to 60-80% of the pool area, the cells were collected according to the nucleoprotein extraction kit (purchased from Shanghai Yi Yan Biotechnology Co., Ltd.) to collect cells and extract cells Protein and nucleoprotein. The interaction of four different  $\beta$ -catenin (X246P) mutants with APC was examined by Flag antibody-mediated Pulldown assay. Specific operation reference [40].

## Results

Cis-peptide bond distribution: from the PDB database file to remove the redundant data to get a 1591 target protein file without a redundant data set, which contains the total peptide key of about 289149. Cis-peptide bond was 854, cis incidence was 0.296% (854/289149), of which 703 were related to proline-related cis-peptide bonds, 82.3% (703/854) of the total cis-peptide bond; Non-proline-related cis-peptide bonds were 151, accounting for 17.68% (151/854) of cis-peptide bonds,

the incidence of cis-peptide bond; B) the percentage of XP and X-XnP cis-peptide bond; C) the incidence of cis-peptide bond in XP peptide bond; D) XP cis-peptide bond with other types of cis Comparison of the number of peptide bonds.

X-P cis-peptide bond incidence: different X-P peptide bond number appears in different, as shown in Table 3.2. As shown in Figure 3.2, the N-terminal residues of the proline peptide bond (incidence > 5%) tyrosine, glycine, tryptophan, phenylalanine, glutamic acid, serine, proline Peptide amide bond has a greater tendency. And methionine, leucine, histidine, isoleucine, valine, glutamine, and lysine have a greater tendency to form trans-peptide bonds.

Table 1 the residues statistics of cis peptide bonds ( X-P)

AP	CP	DP	EP	FP	GP	HP	IP	KP	LP
33	14	31	58	52	73	14	26	31	43
MP	NP	PP	QP	RP	SP	TP	VP	WP	YP
4	32	49	17	28	67	24	31	16	61

The results of Z-fractional analysis showed that when the sequence of XP peptide bond (-5, +5) was rich in glycine, leucine, proline, serine, propylamine Acid, glutamic acid and valine, it is advantageous for XP to form cis-peptide bond. When the current sequence is rich in cysteine, histidine, methionine, tryptophan and asparagine, Conducive to XP cis-peptide bond occurs.

Table 2 Z-scores of residues around cis peptide bond (X- P)

The relative position of the residue to the cis-peptide bond (the cis-peptide bond is formed from the residue at position 1 and the residue at position 1)									
Resid	-5	-4	-3	-2	-1	2	3	4	5
A	1.0546	0.9382	0.3956	1.0278	-0.1143	1.4452	-0.2609	1.0619	0.2297
C	-1.1691	-1.4155	-0.5087	-1.2429	-1.1241	-1.3010	-0.7501	-0.9164	-1.4844
D	0.5829	-0.0425	-0.1319	-0.7051	-0.2206	-0.2025	0.2283	0.0145	-0.1131
E	0.7850	0.4152	0.2449	0.5497	1.2144	0.2781	-0.2609	-0.1018	-0.0446
F	-0.4953	-0.3040	-0.2072	-0.1076	0.8955	1.2392	-0.6958	-0.3927	-0.2503
G	0.3133	0.0229	0.7724	1.6253	2.0116	0.1407	1.5328	-0.1018	0.3668
H	-1.3713	-0.9578	-0.9608	-1.0636	-1.1241	-0.2712	-1.2393	-0.7418	-0.5931
I	-0.0236	0.6767	-0.8101	-0.8844	-0.5394	-1.0264	0.7175	0.1891	0.2297
K	0.1786	0.2844	0.0942	0.3107	-0.2206	-0.4772	0.1196	-0.3927	0.2982
L	1.8632	1.6574	1.6767	1.3266	0.4172	1.7198	0.9349	2.5165	1.7381
M	-1.4387	-1.6771	-1.7898	-1.1234	-1.6555	-1.8503	-1.4567	-1.5564	-1.6901
N	-0.2257	-1.2194	-0.7348	-0.8246	-0.1674	-0.4085	-0.1522	-0.9164	-0.8673
P	-0.4953	0.9382	1.6014	0.7290	0.7361	1.8572	2.6743	1.0619	1.4638
Q	-0.6974	-0.4348	-0.0565	0.0717	-0.9646	0.1407	-1.0762	-0.6255	-0.3188
R	-0.4279	0.2844	0.0188	0.0717	-0.3800	-0.1339	0.3370	-0.2182	0.0926
S	0.5155	0.4152	2.1289	1.3863	1.6927	0.7587	0.8806	1.0037	0.5725
T	0.5829	0.2190	0.2449	0.7290	-0.5926	-0.1339	0.0109	-0.3346	0.2297
V	2.3349	2.1805	0.4710	0.9680	-0.2206	0.0721	0.3370	1.6437	2.1494
W	-0.8996	-1.3501	-1.6391	-1.7807	-1.0178	-1.5757	-1.2937	-1.4401	-1.7586
Y	-0.9670	-0.6309	-0.8101	-1.0636	1.3739	-0.2712	-0.5870	0.2473	-0.2503

(X-XnP): The results showed that X-XnP cis peptides rarely occurred, and only 151 were distributed in 102 different peptides. The highest incidence of X-XnP cis-peptide bond (XG and GX) was 0.173% and 0.156%, respectively, while the tryptophan-associated XW cis-peptide bond also had a high incidence of 0.118 In addition, the incidence of cis-peptide bond of PW, WA, II and GG was more than 0.05%, and the distribution of cis-peptide bonds was found in 278 X-XnP structures such as AD, AE, CA and CD. See Table 3.4.

In order to explore the effect of different XP cis-peptide bond on protein function in eukaryotic cell environment, we constructed four different lentiviral vectors of different  $\beta$ -catenin (X246P) mutants by DNA recombination and point mutagenesis, Where X246P is YP, WP, and LP and MP with the highest incidence of cis-peptide bond with the highest cis-peptide bond

## Discussion

The different definitions of the concept on the cis-peptide bond results: the concept of different cis-peptide bond definition of PDB from the cis-peptide bond rate of the statistical results, the earliest study that the protein multi-titanium chain composition of peptide The 6-related groups of the key are in the same plane, the peptide bond with the dihedral angle equal to  $0^\circ$  is defined as the cis-peptide bond, and the  $\omega$  dihedral angle is equal to  $180^\circ$ , [22]; later studies have found that the

value of  $\omega$  is equal to  $0^\circ$  and  $180^\circ$  rarely occur, so in the PDB will be equal to  $0 \pm 30^\circ$  is defined as cis-peptide bond,  $\omega$  value is equal to  $150 \pm 30^\circ$  definition For trans-peptide bonds [27,29]. We found that the  $\omega$  values of the many peptide bonds were distributed at  $30^\circ < \omega < 90^\circ$  and  $90^\circ < \omega < 180^\circ$  except that the values of  $\omega$  were distributed in the range of  $0 \pm 30^\circ$  and  $150 \pm 30^\circ$ ,  $150^\circ$  range. This means that according to the cis-alien naming rule, many cis-peptide bonds are not counted into the PDP's CISPEP file. Therefore, in this study, we define the cis-peptide bond as  $0^\circ < \omega < 90^\circ$ , and define the trans-peptide bond to a range of  $90^\circ < \omega < 180^\circ$ .

The relationship between the proline structure and the cis-peptide bond: Peptide bonds in the peptide chain are mostly trans-conformation, only a very small number of peptide bonds were cis-conformation, and mostly containing proline peptide, The main reason is that the steric hindrance caused by the tetrahydropyrrole ring in proline eliminates the advantage of trans-conformation, so that the peptide bond formed by proline is relatively easy to cross the barrier between the two conformations to form cis Conformation [30]. Compared with the other 19 amino acid residues, proline is the most specific amino acid [31]. These properties of proline determine that X-P is more likely to form cis-peptide bonds [12, 32, 33] compared to other X-XnP. Thus, the proline-dependent peptide bond in the cis-peptide bond accounted for 82.3% of this study.

Relationship between cis-peptide bond and X-P sequence characteristics: It has been suggested that the occurrence of cis-peptide bonds is related to the presence of proline, but also to the X-P sequence characteristics [10,34]. From the results of this study (see Table 3.3), it can be seen that there is a strong statistical correlation between certain amino acids and cis-peptide bonds at specific positions. When the XP peptide bond is rich in glycine, leucine, proline, serine, alanine, glutamic acid and valine in the front and back sequences (-5, +5), it is advantageous for XP to form cis peptide.

Analysis of the occurrence of non-proline cis-peptide bonds: From the results (see Table 3.4), it can be seen that the probability of occurrence of GX and XG cis-peptide bonds accounts for 39.7% (60% 151), and the highest incidence was GX (0.131%), XG (0.152%). The reason may be that glycine molecules in the same time with acidic and alkaline functional groups, in water can be ionized, has a strong hydrophilic, are polar amino acids, soluble in polar solvents, and difficult to dissolve in non-polar Solvent, and has a high boiling point and melting point, through the aqueous solution of acid and alkaline regulation can make glycine showing different molecular morphology. Some studies have shown that when glycine is involved in the proton transfer process, the phenomenon of reverse energy barrier disappears, which is similar to the hydrogen bond function [10]. Therefore, we speculate that the physicochemical properties of glycine are the cause of this result.

Eukaryotic intracellular studies of the effects of different cis-peptide bond yields on protein function: Since all data in the PDB are derived from the recombinant protein of the prokaryotic expression system, in order to further investigate the cytokine function of eukaryotic cells in order to further investigate the incidence of cis-peptide We investigated the effect of  $\beta$ -catenin on the interaction of  $\beta$ -catenin with APC protein in hepG2 cell line, and the change of cis-peptide bond in S246P site of  $\beta$ -catenin protein was observed. Since it has been shown that  $\beta$ -catenin does not interact with APC molecules when the S246P site is a cis-conformation,  $\beta$ -catenin can not be hydrolyzed and accumulated in the nucleus [35, 36, 37, 38, 39]. Thus, the difference in the incidence of different X-P cis-peptide bonds can be reflected by the intensity of the interaction between  $\beta$ -catenin and APC molecules. The results of this study show that the Y246P and W246P sites represent the highest incidence of cis-peptide bond X-P, whereas the L246P and M246P sites represent the lowest cis-peptide bond rate, X-P.

## Conclusion

The protein cis peptide peptide occurs most often in the XP residue fragment of the polypeptide chain. When X is tyrosine, tryptophan, phenylalanine and glycine, XP has the highest cis peptide bond And X-XnP residues on the X-XnP residue fragment, when the X-XnP is AD, AE, CA, CD and other 278 residues fragments, with the most stable trans peptide bond structure.

## Acknowledgements

This study was supported by the National Natural Science Foundation of China (NSFC: 31501100).

## References

- [1] Zeng Yanfei, Chen Hanchun. Pin1 and cell cycle regulation and related human diseases[J]. Foreign Medicine • Physiology, Pathology and Clinical Volume. 2003, 23 (4)
- [2] Chang Hong. Research and progress of gene diagnosis method[J]. Chinese Journal of Tissue Engineering Research and Clinical Rehabilitation, 2008, 12 (37)
- [3] Chen Yang, Yao Yandan, Luo Manli et al. Progress of prolyl cis-trans isomerase Pin1 in breast cancer [J]. Chinese Journal of Breast Cancer, 2016,10 (5): 300-304.
- [4] Feng Juncai. Organic Chemistry, College of Chemistry and Chemical Engineering, Nanjing University[J]. Science Press, 2012: 82-83.
- [5] Guo Huacan. Advances in post-translational modification of protein[J]. Biotechnology Bulletin, 2011
- [6] Guo Yanting, Li Yanmei, Zhu Zhentai, Zhao Yufen. Effect of phosphorylation on the cis-content of pro-phthalamide bond [J] .Chinese National Chemical Society Symposium on Phosphorus Chemistry and Chemical Engineering, 2003
- [7] Jiang Zheng, Wang Fang, He Xiang et al. Advances in protein phosphorylation modification [J]. Beijing: Biotechnology News, 2009.
- [8] Lai Dongbo et al, hemophilia gene diagnosis research [J]. Chinese Journal of Pediatrics and Cancer, 2006,21 (5): 237-243.
- [9] Li Jun, Du Xin, etc. Progress in protein glycosylation modification[J]. Chinese Science Bulletin, 2009,25 (6): 773-783.
- [10] Li Ping. Proton transfer mechanism and related properties of glycinamide and its derivatives related to cis-peptide bond units [D]. Shandong University, 2006
- [11] Li Weijiang, Zhang Ying. Statistical analysis of the association between proline cis-peptide bond and amino acid sequence[D]. Journal of Inner Mongolia University, 2001
- [12] Li Weijiang, Zhang Ying. Preparation and general characteristics of protein structure model library [J] .Inputational Physics and Theoretical Biology, Inner Mongolia University, Journal of Inner Mongolia University, 2009.
- [13] Li Weijiang, Zhang Ying. Statistical analysis of the association between proline cis-peptide bond and amino acid sequence[D]. Journal of Inner Mongolia University, 2001
- [14] Li Xiang, Zhang Feixiong.Study on histone methylation [J] .However, Beijing: 2004,26 (2): 244-248.
- [15] Liu Keren, Jin Meifang, Wu Shiliang. Prediction of glycosylation sites and catalysis of glycosylation enzymes[J]. Science of Life, 2006, 26 (1).
- [16] Lin Meiju, Yu Yong. Pin1 and tumor [J]. Chinese Journal of General Surgery, 2006,15 (4): 286-288.
- [17] Allen FH, Bellard S, Brice MD, et al. The Cambridge crystallographic data center: computer-based search, retrieval, analysis and display of information [J]. Acta Crystallographica Section B, 1979, 35: 2331-2339 The
- [18] Wang Xiaoyu. ShRNA LKB1 lentiviral vector construction, transfection and expression of[D]. Nanchang University First Affiliated Hospital, 2015.

- [19] Tang Mengxuan, Zhou Wanjun, Hu Yuanjia, etc. HepG2 cell culture methods and conditions. *Practical Preventive Medicine*, 2005, 1: 71-73.
- [20] Andreas Jabs, Manfred S. Weiss and Rolf Hilgenfeld. Non-proline Cis Peptide Bonds in Proteins[J]. *J. Mol. Biol.*, 1999, 286: 291-304.
- [21] IUPAC, *J. Mol. Biol* [J]. 1970, 52: 1-17.
- [22] Wang Minglei, Li Weijiang, Xu Wenbo. Construction of proline skin bond data set [J]. *Computer & Applied Chemistry*, 2003,6 (20).
- [23] <http://www.rcsb.org/pdb/home/home.do>.
- [24] Pierrick Craveur, Agnel Praveen Joseph, Pierre Poulain, Alexandre G. de Brevern, Joseph Rebehmed. Cis-trans isomerization of omega dihedrals in proteins [J]. *Amino Acids*, 2013, 45: 279-289.
- [25] D. Whitford, Wei Qun. *Protein Structure and Function* [M]. Beijing: Science Press, 2008, 14-19.
- [26] Schomaker, V. & Trueblood, K. N. *Acta Cryst*, 1968, B24, 63-76.
- [27] Wang Minglei. *Bioinformatics and molecular dynamics of proline in peptide chain* [D]. Jiangnan University, 2005.
- [28] Andreas Jabs, Manfred S. Weiss and Rolf Hilgenfeld. Non-proline Cis Peptide Bonds in Proteins [J]. *J. Mol. Biol.*, 1999, 286: 291-304.
- [29] Wang Minglei, Li Weijiang. Screening of proline cis-peptide bond by neural network [J]. *Journal of Anhui Agricultural University*, 2003,30 (4): 446-450.
- [30] To Ye. *Natural Protein Toxin Study: Structure and Function of Plant Antifungal Protein EAFP2 and Scorpion Neurotoxin BmK M1* [D]. Institute of Biophysics, Chinese Academy of Sciences, 2003.
- [31] Ye Xinsheng, Shen Pei Fen, Tang Xifang, etc. *Cell exploration regulation*. Beijing: Military Medical Science Press, 1994: 12-15.
- [32] ZHANG Wei. *New Technology for Qualitative and Quantitative Analysis of N-Glycosylation Based on Biomass Mass Spectrometry* [D]. Fudan University, 2012.
- [33] ZHU Zhen-tai. *Study on the effect of phosphorylation on the cis-trans isomerization of peptide prolyl amines* [D]. Beijing: Department of Chemistry, Tsinghua University, 2005.
- [34] Allen FH, Bellard S, Brice MD, et al .. The Cambridge crystallographic data center: computer-based search, retrieval, analysis and display of information [J]. *Acta Crystallographica Section B*, 1979, 35: 2331-2339
- [35] Shin HR, Islam R, Yoon WJ, Lee T, Cho YD, Bae HS, Kim BS, Woo KM, Baek JH, Ryoo HM. Pin1-mediated protein prolonged the nuclear retention of  $\beta$ -catenin in Wnt3a-induced osteoblast Differentiation [J]. *J Biol Chem*. 2016 Jan 6. pii: jbc.M115.698563
- [36] Natoli M, Leoni BD, D Agnano I, D'Onofrio M, Brandi R, Arisi I, Zucco F, Felsani A. Cell growing density affects the structural and functional properties of Caco-2 differentiated monolayer [J] *Cell Physiol*. 2011, 226 (6): 1531-43.
- [37] Sambuy Y, De Angelis I, Ranaldi G, Scarino ML, Stammati A, Zucco F. The Caco-2 cell line as a model of the main barrier: influence of cell and culture-related factors on Caco-2 cell functional Characteristics [J]. *Cell Biol Toxicol*. 2005, 21 (1): 1-26.
- [38] So KY, Oh SH. Prolyl isomerase Pin1 regulates cadmium-induced autophagy via ubiquitin-mediated post-translational stabilization of phospho-Ser GSK3 $\alpha$  $\beta$  in human hepatocellular carcinoma cells [J]. *Biochem Pharmacol*. 2015, 98 (3): 511 -521.

- [39] Ryo A, Nakamura M, Wulf G, Liou YC, Lu KP. Pin1 regulates turnover and subcellular localization of beta-catenin by adhere its interaction with APC. [J]. Nat Cell Biol. 2001, 3 (9): 793-801.
- [40] Anna P, Teresa P. Protein-protein interactions and networks [M]. Maryland: National Institutes of health bethesda Press, 2008.
- [41] Yang Yingbin. Bone marrow stem cell-derived malignant transformation of target cell clearance [J] .College of Bioengineering, Chongqing University, 2010