TPR Packing Analysis and 3D Modeling for the HAT Domain of Human Crooked Neck Protein

Zhenglin Hou, Cunxi Wang, Odd-Arne Olsen

Pioneer Hi-bred International Inc.
Johnston, Iowa 50131-0552
Zhenglin.Hou@pioneer.com

Abstract

Human crooked neck protein (hcrn) containing 17 HAT or TPR repeats plays a role in pre-mRNA processing. Conserved residues in the TPR consensus sequence of 34 aa were found at helical packing interface and pro32 which breaks the second helix. The crn TPR helical hairpins were built on consensus TPR 3d template and packed side by side to form the overall superhelical structure. The models underwent a series of energy minimizing refinements and molecular dynamics simulations under constrains of holding each helical structure together but allow individual helix to spin around its own axis. The refined structures preserved the main characteristics of TPR superhelical fold with every 7 TPR units forming a complete repeat. The knob-hole rule was satisfied in majority of helix-helix packing. The models indicated that hcrn exerts its function in either mRNA processing or DNA duplication by mediating protein-protein interaction in a complex assembly.

1. Introduction

The Human crooked neck protein (hcrn) is a component of spliceosome and plays an essential role in pre-mRNA processing as extracts depleted of this protein fail to splice pre-mRNA [1]. The spliceosome is a large ribonuceoprotein complex assembled from 5 small nuclear RNAs and more than 50 proteins. The hcrn protein composed almost entirely of 17 TPR (tetratricopeptide) repeats was speculated to serve as a scaffold organizing the multi-protein complex. TPR-containing proteins have been found in wide spectrum of species and participate in diverse biological functions, ranging from control of transcription initiation to RNA processing to protein folding, modification, and proteolysis [2]. The 34 aa repeated motifs are highly divergent in sequence with only a small number of conserved residues while their three 3d structures show a remarkable similarity with two α-helices forming a tightly packed antiparallel hairpin. Often multiple TPR hairpins stacked together in a parallel array to produce an overall superhelical architecture and create an extended groove suitable for legand binding. Many TPR crystal structures have been solved but these structures all contain a few TPR repeats (<4) due to crystallization limitation. Here we reported a structural model of hcrn of 17 TPR repeats.

2. Results and Discussions

TPR repeats appear to possess high amino acid substitution rates and thus recognition of repeat homologues is highly problematic. The hcrn TPRs were identified based on multiple tools including HMM search against Pfam, Superfamily database, REP [3, 4], and GCG. 17 TPRs were found after manual consolidation. Historically, the hcrn repeated motifs have been called "half a TPR" (HAT) repeats due in part to the low conservation on a few positions of small amino acids. However, along with the sequence accumulation, the distinction between HAT and TPR has become obscure and the HAT is an actual TPR with a slight different repeat unit definition. The identified TPRs were aligned together based on profile comparison with an emphasis on the conserved positions (Figure 1).

```
TPR1   208   PPPQQRITDEEELN-DYKLRKRKTFEDNIRKNRTV
TPR2   242   ISNWIKYAQWEESL-KEIQRARSIYERALDVDYRN
TPR3   276   ITLWLKYAEMEMKN-RQVNHARNIWDRAITTLPRV
TPR4   310   NQFWYKYTYMEEML-GNVAGARQVFERWMEWQPE
TPR5   343   EQAWHSYINFELRY-KEVDRARTIYERFVLVHPD
TPR6   376   VKNWIKYARFEEKH-AYFAHARKVYERAVEFFGDEHMD
TPR7   413   EHLYVAFAKFEENQ-KEFERVRVIYKYALDRISKQDA
TPR8   449   QELFKNYTIFEKKF-IIVSKRRFQYEEEVKANPHN
TPR9   491   YDAWFDYLRLVESD/AEAEAVREVYERAIANVPPIQ/
TPR10  534   IYLWINYALYEELEaKDPERTRQVYQASLELIPHKKFTF
TPR11  573   AKMWILYAQFEIRQ-KNLSLARRALGTSIGKCEK
TPR12  606   NKLFKVYIELELQL-REFDRCRKLYEKFLEFGPEN
TPR13  640   CTSWIKFAELETIL-GDIDRARAIYELAISQPRLDMP
TPR14  676   EVLWKSYIDFEIEQ-EETERIRNLYRRLLQRTQH
TPR15  709   VKVWISFAQFELSS/GSLTKCRQIYEEANKTMRNCEE/
TPR16  755   LESWRSFEEEFGTA-SDKERVDKLMPEKVKKRRKVQTDDGS
TPR17  795   DAGWEEYEDYLIFPE-DAANQENLKLLAMAKLWKKQQ
```

Figure 1: hcrn TPRs alignment.

The hcrn sequence profile matched best to an idealized TPR CTPR3 (PDB code 1na0) among all known TPR structures. CTPR3 contains three identical TPRs and the TPR motif was designed according to the consensus sequence of naturally occurred TPRs. The structural template was constructed by stacking

multiple cTPRs together following the packing parameters on average. The structural models were built using MOLERLER [5], in which the calculated structural models were obtained by optimally satisfying spatial restraints derived from sequence and structure alignment. The raw model was underwent a series of energy minimizing refinements and molecular dynamics simulations under constrains of holding each helical structure together but allow individual helix to spin around its own axis.

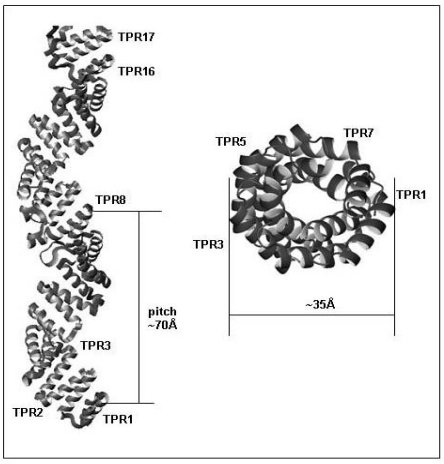

Figure 2. The overall structure

The overall hcrn shows a superhelical architecture with more two repeats (Figure 2). Each repeat is roughly 70 Å and contains seven TPR units which stacking together with packing angles, A-B helices ~-162^{o}, B-A' ~-150^{o}, and A-A' ~31^{o}. The diameter of the superhelical structure is about ~35 Å.

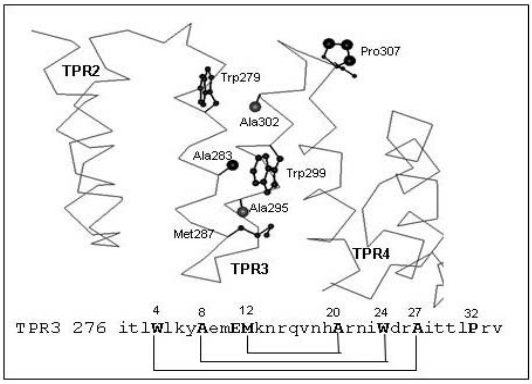

Figure 4. Conserved helix-helix packing interface.

The refined structures preserved the main characteristics of TPR fold and only a few extended loops linking helices have a big uncertainty. The first and last TPR units, less conserved than others, might not adopt the typical TPR conformations but probably stabilize the overall structure. The inter-helical packing roughly follows the knob-hole rule, i.e. large residues in one helix stack with small resides on the other helix (Figure 3).

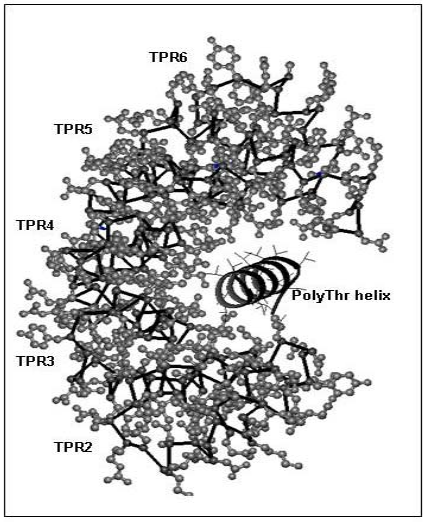

Figure 3. Putative helix binding in the groove

The extended groove mainly made of hydrophilic residues has potential to bind helical peptides, and the long superhelix could simultaneously interact with multiple proteins (figure 4). Presumably, hcrn mediate the protein-protein interaction in spliceosome and glue the complex together.

3. References

[1] S. Chung, Z. Zhou, K.A. Huddleston, D.A. Harrison, R. Reed, T.A. Coleman, B.C. Rymond, "Crooked neck is a component of the human spliceosome and implicated in the splicing process", Biochim Biophys Acta. 2002; 1576:287-97

[2] L.D. D'Andrea, L. Regan, "TPR proteins: the versatile helix", Trends Biochem Sci. 2003; 28:655-62

[3] S.R. Eddy, Profile hidden Markov models. "Bioinformatics", 1998; 14; 755–763.

[4] M.A. Andrade, C.P. Ponting, T.J. Gibson, P. Bork, "Homology-based method for identification of protein repeats using statistical significance estimates". *J Mol Biol.* (2000), **298**;521-537

[5] A. Sali and T. Blundell, "Comparative protein modeling by satisfaction of spatial restraints", *J. Mol. Biol.,* 1993; **234**:779-815