Crystal structure determination of a neutral neurotoxin BmK M4 from *Buthus martensii* Karsch at 0.20 nm

HE Xiaolin, LIU Xinqi, ZENG Zonghao,
LI Hongmin, WANG Miao, ZHANG Ying
& WANG Dacheng

Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
Correspondence should be addressed to Wang Dacheng (email: wdc@pewdc.ibt.ac.cn)

Received May 19, 1999

Abstract BmK M4 is a neutral neurotoxin in the BmK toxin series. It is medially toxic and belongs to group III α-toxins. The purified sample was crystallized in rhombic space group $P6_1$. Using an X-ray diffraction technique, the crystal structure of BmK M4 was revealed by molecular replacement at 0.20 nm resolution. The model was refined. The final crystallographic *R* factor was 0.142 and the free *R* factor was 0.173. The root mean square deviation is 0.001 5 nm for the bond length and 1.753° for the bond angles. 64 water molecules were added to the asymmetric unit. The refined structure showed an unusual non-prolyl *cis* peptide bond at residue 10. The structure was compared with group II α-toxin BmK M8 (an acidic, weak toxin). The potential structural implications of the *cis* peptide bond were discussed.

Keywords: scorpion neurotoxin, neutral toxin molecule, crystal structure, *cis* peptide bond.

Scorpions are widely distributed in the world, attacking human and animals with its tail, causing serious symptoms. These symptoms are caused by the interaction of multiple active peptides and proteins from the tail-gland venom with the excitable membrane in the animal neural-muscular system. These are neurotoxins. It is known that scorpion neurotoxins are divided into short-chained and long-chained families. The short-chained toxins have 30—40 residues, 3—4 disulfide bridges. The long-chained toxins have 60—70 residues and are crosslinked by 4 disulfide bonds. They selectively bind to voltage-dependent sodium channel and affect the sodium permeability in the excitable cells[1]. These toxins vary in homology sequence and can be divided into mammalian, insect and crustacean selective toxins according to the specificity of different animals. The mammalian toxin can be further divided into α- and β-toxin according to their different influence on sodium current and different receptor-binding properties[2]. The α-toxins delay the inactivation of sodium current and bind to receptor site 3 on sodium channel; whereas the β-toxin inhibits the activation of the sodium current and binds to the receptor site 4. Furthermore,

Abbreviations: BmK, East-Asia scorpion *Buthus martensii* Karsch; AaH, North-Africa scorpion *Androctonus australis* Hector; CsE, North-America scorpion *Centruroides sculpturatus* Ewing.

the α-toxins have been classified into groups I, II, III, IV[3] according to their sequential and

immunological properties. Since there are kinds of scorpion toxins with various functions, they are preferable materials for detailed analysis of the protein structure/function relationship and membrane ligand-receptor interaction.

Buthus martensii Karsch (BmK) is widely distributed in East-Asia and China. It has unique biological toxicity and pharmacological properties. Scorpions from some other regions in the world (e.g. AaH in the North Africa) can cause death of human or animals, and thus form an environmental-medical problem, whereas BmK is comparatively less toxic and has never caused the death of a human. On the contrary, the whole scorpion (especially the tail) has been used in the treatment of neural diseases for more than 1 000 years. So the study of BmK has a unique biological and medical background.

Recently, some components with significant pharmacological effects have been purified from BmK venom and their genes cloned[4], with a bioactivity-variant series of neurotoxins purified and crystallized[5], and also cloned into vectors, then overexpressed[6,7]. These were valuable for the study of structure-function relationship. Significantly, in this series (BmK M1, M2, M4 and M8), they vary in activity (20 times as active), and belong to different α-toxin groups. BmK M1, M2, M4 belong to group III[8], whereas BmK M8 belongs to group II[9], because of different binding specificity. Previously, we have reported the crystal structure of the acidic toxin BmK M8[9,10]. With the addition of the structure of the potent toxin AaH II from North-Africa scorpion[11], only two high-resolution crystal structures of scorpion α-toxins have been determined. This study reports the structure and analysis of a neural medially toxic component BmK M4 at 0.20 nm resolution. Interestingly, the BmK M4 structure revealed a non-prolyl *cis* peptide bond that was first found in scorpion toxin which is also rare in general protein structures.

1 Materials and methods

1.1 Purification and crystallization

BmK M4 was prepared from scorpion venom with several cycles of chromatography as described previously[5]. For high purity, the samples were further purified by preparative isoelectric focusing using polybuffering ampholine (pH 7—10). The purity was more than 98%, as judged by analytical IEF and HPLC analysis. BmK M4 has 64 residues, with MW of 7001. The sequence is VRDGYIAKPH NCVYECARNE YCNDLCTKDG AKSGYCQWVG KYGNGCWCIE LPDNV PIRVP GKCH. Its pI is 7.53. Tests on animals found it medically toxic to rats (half-lethal dose 40 mg/kg). It was also weakly toxic to fly larva (half contraction dose 3.0 mg/kg).

The crystals were grown as previously[5] described by hanging-drop vapor diffusion method. The protein was dissolved with 1 mmol/L HAc into 20 mg/mL. The well solution was 48 mol/L ammonium acetate. 5 μL protein was mixed with 5 μL well solution and then equilibrated against 1 mL well solution. Crystals appeared after one week and grew to 0.6 mm, its maximum dimension, in one month. The crystal has obvious 6-fold axis. The cell dimensions determined on an IP

detector were $a=b=5.558$ nm, $c=3.39$ nm, $a=b=90^\circ$, $g=120^\circ$. The space group was $P6_1$ or $P6_5$. In Mathews calculation, assuming one protein molecule per asymmetric unit, the Vm value was 0.002 07 nm^3/dalton and the solvent content was 40.6%, falling into reasonable region.

1.2 Diffraction data collection

The diffraction data were collected with a Siemens X-200B area detector at room temperature. The monochromatized CuKα radiation was generated by a Rigaku RU200 rotatory anode(40 kW, 200 mA). All data were from one crystal. The resolution was better than 0.20 nm. The evaluation, scaling, and merging were accomplished by XENGEN. In the range of 2.93—0.20 nm, 4 091 unique reflections were obtained from 11 618 measurements, with a completeness of 98.9% and R_{merge} of 2.79% ($R_{merge}=\Sigma(I(i)-\langle I\rangle)/\Sigma I(i)$. $I(i)$, measured intensity for every reflection; $\langle I\rangle$, the average intensity of multiple measured reflections), showing that it is a high-quality data set.

1.3 Structure determination and refinement

The structure of BmK M4 was solved by molecular replacement with AmoRe in CCP4 package. We have determined the crystal structures of the acidic toxin BmK M8 [9] and a basic toxin BmK M1 (to be published). Sequence comparison[5] shows that BmK M4 is closer to BmK M4. So BmK M1 served as the probe. No substitution was applied to the different residues. Using the $I/\sigma(I)\geqslant 2$ data, the rotation function and translation function were calculated. The potential solution was subjected to rigid-body refinement and checking of crystal packing with TURBO-FRODO.

The initial model obtained from molecular replacement was refined. Firstly, the correct residues were recovered with TURBO-FRODO. Then conjugate gradient minimization and slow cooling were applied using the calculated phase(a_c), and amplitude (F_c) and amplitude (F_o) were observed. The$(2F_o\text{-}F_c)$, $(F_o\text{-}F_c)$, and omit$(F_o\text{-}F_c)$ electron density map was calculated. The model was rebuilt with TURBO-FRODO. Then energy minimization restrained least square positional refinement and thermal factor refinements were applied. A new cycle of model rebuilding was based on the newly refined coordinates. The crystallographic refinement and model building were repeated until convergence was reached. Last, the water molecules were added. The refined model was analyzed with TURBO-FRODO and PROCHECK.

2 Result and discussion

2.1 Molecular replacement solution of BmK M4

Based on the structure of BmK M1, rotation functions were searched for different resolution ranges. The integration radius was 0.18 nm. The step of a, b and γ was 3° (a, b and γ, the Euler angles defined by Growther). Consequently, when adopting a resolution range of 1.0—0.4 nm, the highest peak with correlation coefficient (CC) of 20.5 was found at $a=37.1^\circ$, $b=112.1^\circ$,

g=181.1°, whereas the second highest peak had a CC only 17.5. The highest peak was 4.8s above the mean.

Translation function was calculated for two possible space groups $P6_1$ and $P6_5$. In resolution range of 1.0—0.35 nm, the translation functions were studied on the basis of the rotation solution. Consequently, when assuming space group $P6_5$, the highest peak had a CC of 23.8 and R factor of 52.2%, the second peak a CC of 22.7 and R factor of 52.8%; whereas when assuming space group $P6_1$, the highest peak had a CC of 34.6 and R factor of 49.7%, the second peak having a CC of 21.5 and R factor of 52.3%. This shows that $P6_1$ is the correct space group. The corresponding best translation was T_x=0.983 3, T_y =0.650 0, T_z=0.000 0. Using the rigid-body refinement within AmoRe, the refined solution a=35.70°, b=111.01°, γ =181.76°, T_x=0.991 0, T_y=0.639 6, T_z=0.000 0 gave a CC of 58.3 and R factor of 0.423. Applying the solution on the probe model and checking of the molecular packing found no conflict. The distance from neighbor molecules was also reasonable. So the correctness of molecular replacement solution was verified.

2.2 Crystallographic refinement and model building

The initial model based on the BmK M1 structure has an R factor of 42.3%. The 12 different residues were directly substituted for correct ones and the resulting side chains had random torsion angles. 10% data were set for random cross validation. The refinement was guided by free R factor. Stereochemical restraints were applied. All $I/\sigma(I)\geqslant 0$ data were used. $2(F_o\text{-}F_c)$ and $(F_o\text{-}F_c)$ electron density maps were calculated to guide the rebuilding of local regions. For regions with poor density, omit $(F_o\text{-}F_c)$ maps were calculated. The detailed process of refinement was tabulated in table 1. The final R factor and R_{free} were 14.2% and 17.3%, respectively. The root mean square deviation of bond length was 0.001 5 nm, whereas that of bond angles was 1.752°. The good fit of the model with the data and stereochemistry indicate that it is a reasonable model.

Table 1 Molecular dynamic refinement process

Stage	R	R_{free}	Resolution range/nm	$RMSD_{bond}$/nm	$RMSD_{angle}$/(°)	Notes
0	0.423	-------	1.0—0.35	0.002 0	2.264	result of molecular replacement
1	0.386	0.456	1.0—0.30	0.001 6	1.984	1 cycle of rigid-body refinement, 2 cycles of positional refinement
2	0.302	0.376	1.0—0.30	0.001 6	1.813	1 cycle of simulated annealing, 1 cycle of positional refinement
3	0.274	0.340	1.0—0.20	0.0016	1.782	1 cycle of simulated annealing, 1 cycle of positional refinement
4	0.204	0.249	1.0—0.20	0.001 5	1.759	alteration between hand modeling and positional, B factor refinement
5	0.142	0.173	2.0—0.20	0.001 5	1.752	adding 64 water molecules, then positional, B factor refinement and hand modeling

2.3 Electron density maps

The main chain of BmK M4 fits well with the final electron density map. For the side chains, only the side chain terminus of Lys24, Glu32, Glu50 and Lys62 have no well-shaped electron density on the ($2F_o$-F_c) map contoured at 1.0s. These four residues are long-chained and on the surface, having flexibility naturally. All the other residues have well-shaped electron density for their side chains. Fig.1 shows the fitting of the functionally important residue Tyr5 to the electron density map. It could be found that the aromatic loop is hollowed, and the water molecules have sphere densities, showing high quality of the refined model. The Ramachandran plot (fig. 2) shows all the main chain conformational angles fall into core and allowed regions, with 86% of them in core regions. So the structure has good main chain stereochemistry.

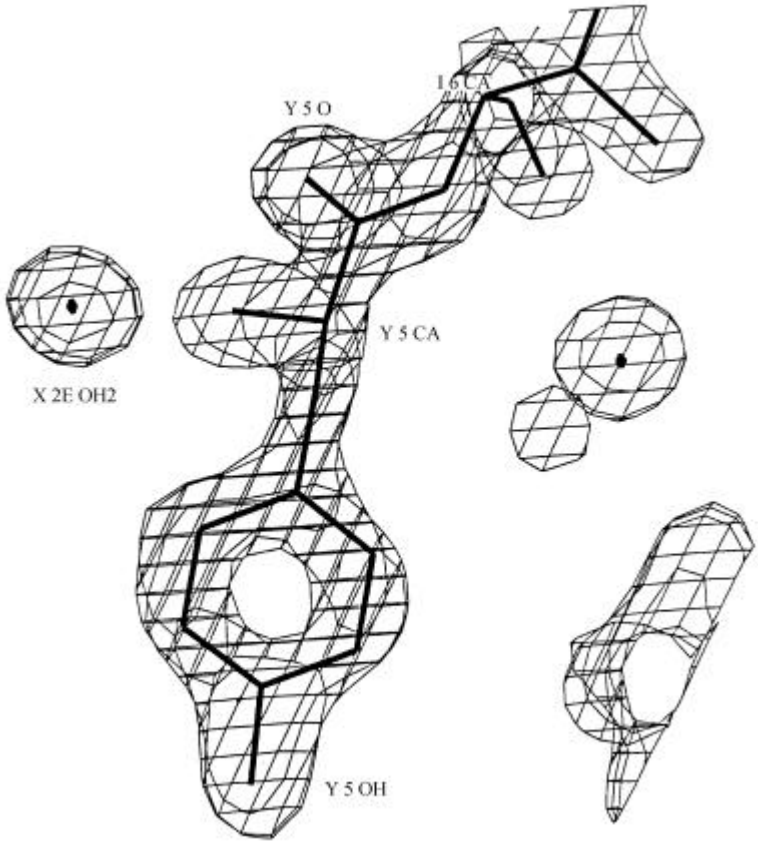

Fig. 1. The $2F_o$-F_c electron density map around Tyr 5 contoured at 1.5 σ.

2.4 Overall structure of BmK M4: common scaffold

The overall structure of BmK M4 was made of an α helix and a three-stranded β-sheet (fig. 3). They form a dense core. The helix (H) is made of residues 19—28; whereas the three β -sheet strands are made of residues 2—5(β I), 31—39(β II), 45—51(β III), respectively. Four disulfide bonds (S1-S4: Cys12-Cys63, Cys16-Cys36, Cys22-Cys46 and Cys26-Cys48) crosslink the secon-

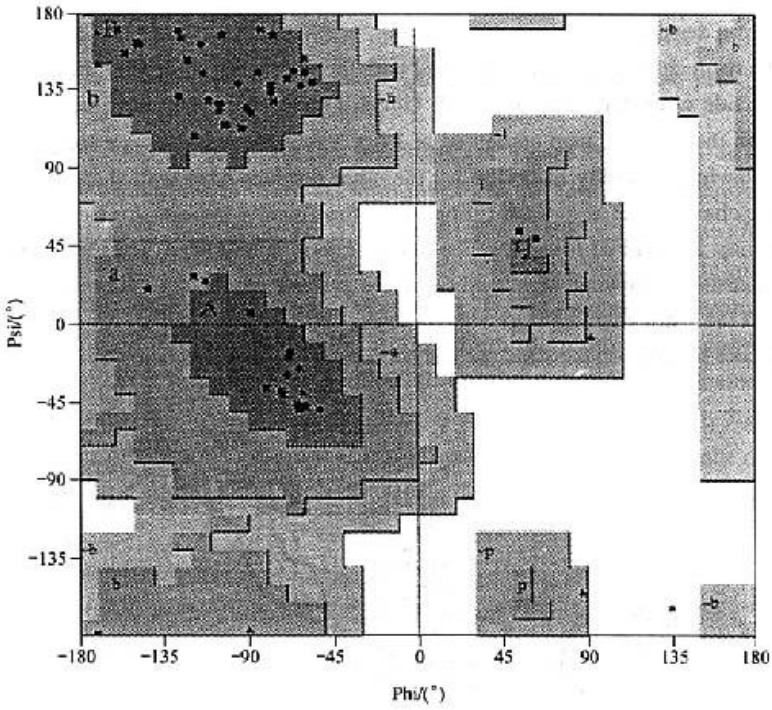

Fig. 2. Ramachandran conformational plot produced by PROCHECK, in which 86% of residues fall into core regions and 14% fall into allowed regions.

dary structure elements to form a disulfide-bridge-locked βαββ motif, which is the scaffold of the structure. This scaffold not only exists in all scorpion toxins, but also exists in many functionally different proteins (e.g., plant and insect defensins[12], sweet protein[13]). As a good basis for protein engineering, this scaffold is receiving more and more attention[14]. Based on the scaffold, three loops including L1 (residues 6—18), L2 (residues 39—44), L3 (residues 52—64), protrude on the molecular surface. The loops can have different conformations in scorpion toxin from different species. They are the easily-changed parts on both sequence and spatial structure, and may concern the specificity and functional difference of different kinds of toxins.

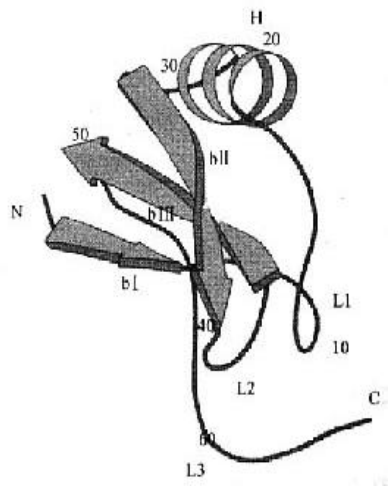

Fig. 3. Ribbons plot of BmK M4, in which the α-helix, β sheet and loops are represented by H, b, and L, respectively.

2.5 Unusual non-prolyl *cis* peptide bond

Since peptide bonds are partially double-bonded, two configurations, *cis* and *trans*, exist. The energy of the *cis* form is much higher than the *trans* form. So in protein structures, peptide bonds are mainly *trans*. It has been reported that only 0.36% of all the peptide bonds are *cis*, with most of those occurring before proline[15]. The rare occasions in which non-prolyl peptide bonds appeared showed it must have functional or structural importance. The refined BmK M4 structure definitely showed that the peptide bond before Glu 10 is *cis* formed (fig. 4). This unusual *cis* peptide bond exists in a reverse turn formed by residues 8=12. Detailed analyses showed that the *cis* peptide bond is correlated with Pro 9. For the special restraint of proline to the main chain structure, the Cβ of residue 10 would be hindered spatially with S_γ of Cys 12 if the peptide bond 9-10 was *trans*-formed. So the 9-10 peptide bond has to be *cis*-formed.

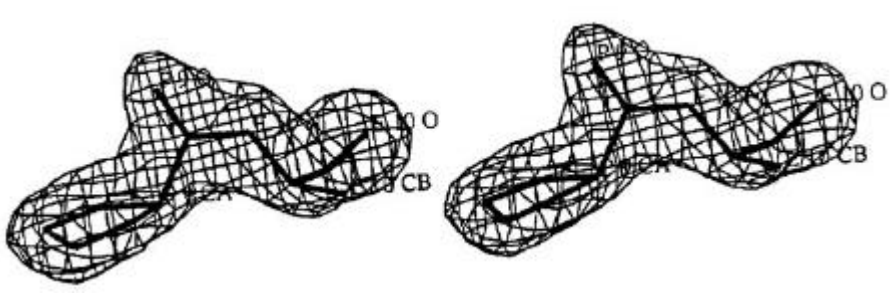

Fig. 4. The omit F_o-F_c electron density map for the main chain of residues 9 and 10, contoured at 2.5 **s**. clearly shows the *cis* configuration of the peptide bond 9-10. Residues 9 and 10 were omitted in the calculation of the map.

2.6 The structural implication of the *cis* peptide bond

BmK M4 and M8 belong to group II and group III α-scorpion toxins, respectively, both structures determined at high resolution. Based on competitive experiments on sodium channel, a recent report [3] proposed that the two groups of toxins may bind to different sites on sodium channel. So finding the structural basis of the binding specificity is very important. Fig. 5 shows the overlapping of BmK M4 and BmK M8 backbone. It could be found that the main difference appears at the C-terminal segment correlated with the reverse turn (8-12). Here BmK M4 has *cis* peptide bond at position 9-10, whereas BmK M8 has a *trans*-formed one. Correspondingly, the C-terminal segment adopts different spatial orientation in these two molecules. In BmK M4, the C-terminus points to segment 8-12. Moreover, the terminal residue His 64 forms two main chain hydrogen bonds (H64 O···E10 N, 0.308 nm; H64 N···E10 O, 0.298 nm) with Glu 10, where the peptide bond is *cis* formed (fig. 6). This suggests that the location of C-terminus is correlated with the *cis* peptide bond, while for BmK M8, the C-terminus is removed from segment 8-12. Obviously, this difference may be of concern to the binding specificity on the receptor.

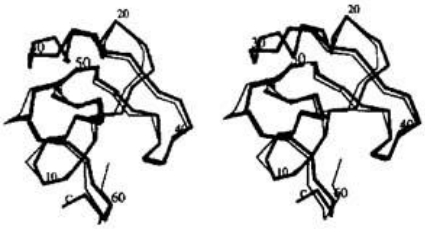

Fig. 5. The stereo overlapping of BmK M4 and BmK M8. The thick line represents BmK M4 and the thin line represents BmK M8.

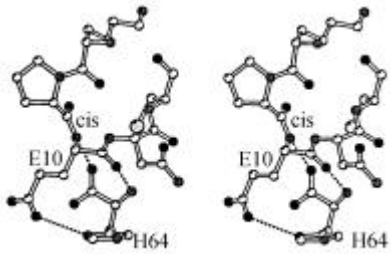

Fig. 6. The hydrogen bonding between the *cis* peptide bond at position 10 and the terminal residue H64 in the structure of BmK M4. The hydrogen bonds are dashed.

Acknowledgements This work was supported by the "863" Project and National Natural Science Foundation of China (Grant No. 39670158).

References

1. Rochat, H., Benard, P., Couraud, F., Scorpion toxins: chemistry and mode of action, in Advances in Cytopharmacology (eds. Cecareli, B., Clementi, F.), New York: Raven Press, 1979, 3: 357.
2. Couraud, F., Jover, E., Debois, J. M. et al., Two types of scorpion toxin receptor sites, one related to activation, the other to the inactivation of the action potential sodium channel, Toxicon, 1982, 20: 9.
3. Gordon, D., Matineauclaire, M. F., Cestele, S. et al., Scorpion toxins affecting sodium current inactivation bind to distinct homologous receptor sites on rat brain and insect sodium channel, J. Biol. Chem., 1996, 271: 8034.
4. Xiong, Y. M., Ling, M. H., Lan, Z. D. et al., The cDNA sequence of an excitatory insect selective neurotoxin from the scorpion *Buthus martensii* Larsch with analgesic activity from scorpion BmK, Toxicon, 1999, 37: 335.
5. Li, H. M., Zhao, T., Jin, L. et al., A series of bioactivity-variant neurotoxins from scorpion *Buthus martensii* Karsch: purification, crystallization and crystallographic analysis, Acta Cryst., 1999, D55.
6. Xiong, Y. M., Ling, M. H., Wang, D. C. et al., The cDNA and genomic DNA sequences of a mammalian neurotoxin from the scorpion *Buthus martensii* Karsch, Toxicon, 1997, 35: 1025.
7. Xiong, Y. M., Ling, M. H., Zhao, T. et al., The cDNA sequences of two anti-mammalian toxins from East-Asia Scorpio, Acta Biochem. Biophys. Sinica, 1997, 29: 200.
8. Luo, M. J., Xiong, Y. M., Wang, M. et al., Purification and sequence determination of a new neutral mammalian neurotoxin from the scorpion *Buthus martensii* Karsch, Toxicon, 1997, 35: 723.
9. Li, H. M., Wang, D. C., Zeng, Z. H. et al., Crystal structure of an acidic neurotoxin from scorpion *Buthus martensii* Karsch at 1.85 Å resolution, J. Mol. Biol., 1996, 261: 415.
10. Li, H. M., Jin, L., Zeng, Z. H. et al., Structure determination of an acidic neurotoxin (BmK M8) from *Buthus martensii* Karsch at 0.25 nm resolution, Science in China (in Chinese), Ser. C, 1996, 26: 213.
11. Fontecilla-Camps, J. C., Habersetzer-Rochat, C., Orthorhomic crystals and three-dimensional structure of the potent toxin II from the scorpion *Androctonus australis* Hector, Proc. Natl. Acad. Sci. USA, 1988, 85: 7443.
12. Broekaer, W. F., Terras, F. R. G., Lammue, B. P. A. et al., Plant defensins: novel antimicrobial peptides as components of the host defense system, Plant Physiol., 1995, 108: 1353.
13. Gao, G. H., Dai, J. X., Ding, M. et al., Solution conformation of brazzin by ^{1}H nuclear magnetic resonance: resonance assignment and secondary structure, Int. J. Biomacromol., 1999, 24: 351.
14. Vita, C., Engineering novel proteins by transfer of active site to natural scaffolds, Curr. Opion. Biotech., 1997, 8: 429.
15. Stewart, D. E., Sarker, A., Wampler, J. E., Occurrence and role of *cis* peptide bonds in protein structures, J. Mol. Biol., 1990, 214: 253.