

# pKNOT: the protein KNOT web server

Yan-Long Lai<sup>1</sup>, Shih-Chung Yen<sup>1</sup>, Sung-Huan Yu<sup>1</sup> and Jenn-Kang Hwang<sup>1,2,\*</sup>

<sup>1</sup>Institute of Bioinformatics, National Chiao Tung University, Hsinchu 30050, Taiwan and <sup>2</sup>Core Facility for Structural Bioinformatics, National Chiao Tung University, Hsinchu 30050, Taiwan

Received January 31, 2007; Revised March 31, 2007; Accepted April 14, 2007

## ABSTRACT

**Knotted proteins are more commonly observed in recent years due to the enormously growing number of structures in the Protein Data Bank (PDB). Studies show that the knot regions contribute to both ligand binding and enzyme activity in proteins such as the chromophore-binding domain of phytochrome, ketol-acid reductoisomerase or SpoU methyltransferase. However, there are still many misidentified knots published in the literature due to the absence of a convenient web tool available to the general biologists. Here, we present the first web server to detect the knots in proteins as well as provide information on knotted proteins in PDB—the protein KNOT (pKNOT) web server. In pKNOT, users can either input PDB ID or upload protein coordinates in the PDB format. The pKNOT web server will detect the knots in the protein using the Taylor’s smoothing algorithm. All the detected knots can be visually inspected using a Java-based 3D graphics viewer. We believe that the pKNOT web server will be useful to both biologists in general and structural biologists in particular.**

## INTRODUCTION

Knotted proteins have become more common in recent years (1–14) due to the enormously growing number of structures deposited in the Protein Data Bank (PDB). The knots in proteins are more than just topological novelties. The knotted regions have been shown to be important in both ligand binding and enzyme activity. For example, the unique knot topology in bacterial phytochrome (6) is common to all red/far-red photochromic phytochrome and is important in stabilizing the chromophore-binding region. The knot regions in TrmD tRNA methyltransferase (MTase) have been shown to be important for S-adenosyl-L-methionine (AdoMet) binding and catalytic activity (3). The deep trefoil knot region in *N*-acetylmethionine transcarbamylase forms part of the

active site (10). The figure-eight knot in the mainly  $\alpha$ -helical domain of ketol-acid reductoisomerase (KARI) forms most of the keto-acid substrate-binding site (11). In addition, knots in proteins present a challenge in the study of protein folding, for it is hard to image a peptide chain to thread through a hoop to form a knot in a reproducible way (15). Interestingly, a recent study (16) showed that YibK (4), a SpoU MTase containing a deep trefoil knot, is able to fold efficiently and behaves remarkably similar to other proteins.

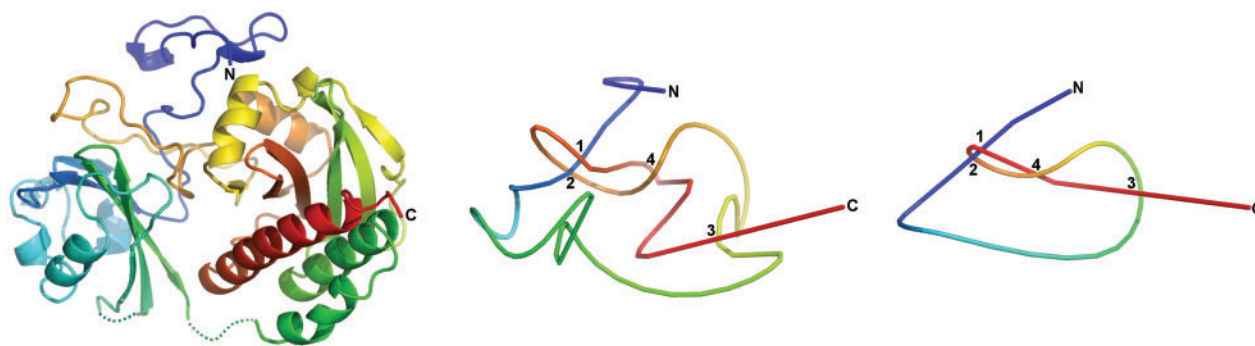
Though the identification of a general knot is a topologically difficult problem, it is relatively easy to identify knots in proteins. However, there were still many cases of misidentified knots in proteins (17,18) due to the lack of a convenient tool available to general biologists. The causes of the misidentification of knots in proteins may be due to the presence of mobile loops, missing residues or just visual error in tracing out the entangled protein chains. For example, the SET domain was originally identified to have a knot, but later it was pointed out that part of the loop relevant to the formation of the knot is in fact connected through hydrogen bonds (17). As a result, the knot in the SET domain turns out not to be an authentic one. Other examples of misidentified knots are the trefoil knot in clathrin D6 coat protein (19), the left-handed trefoil knot in ubiquitin hydrolase (15) and the figure-eight knot in histone K79 methyltransferase (19). These knots are in fact caused by breaks in the chain and are therefore not authentic knots. A more recent example is the misidentified trefoil knot in the chromophore-binding domain of phytochrome (6), which in fact contains a figure-eight knot.

## METHOD AND IMPLEMENTATION

The pKNOT web server detects the knot in a protein by smoothing the protein chain using the Taylor’s algorithm (15). The algorithm first fixes both N and C termini in space, then repeatedly smoothes and straightens the protein chain. The chain is reduced in such a way that, with details of the chains eliminated, the knot can be easily detected. If the protein does not

\*To whom correspondence should be addressed. Tel: +886-3-513-1337; Fax: +886-3-572-9288; Email: jkhwang@cc.nctu.edu.tw

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors



**Figure 1.** An example of the chain smoothing process. The protein is the chromophore-binding domain of bacterial phytochrome (1ZTU). The X-ray structure of 1ZTU (left) is shown in the cartoon representation and two progressively smoothed chains are shown (center and right). The color is ramped by residues from blue at the N-terminus (labeled by N) to red at the C-terminus (labeled by C). The crossover points are numbered sequentially from the N-terminus. The figure-eight knot is characterized by four crossover points, alternately under and over. The structural pictures are produced using Pymol (Delano Scientific, San Carlos, <http://pymol.sourceforge.net/>).

contain a knot, the chain will simply shrink into a straight line. The Taylor's algorithm formally goes as follows: Let the protein chain of length  $N$  be described by  $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ , where  $\mathbf{r}_i$  is the coordinate of the  $i$ -th  $C_\alpha$  atom. A new coordinate  $\mathbf{r}'_i$  is taken to be  $\mathbf{r}'_i = (\mathbf{r}_{i-1} + \mathbf{r}_i + \mathbf{r}_{i+1})/3$ , where  $2 \leq i \leq N-1$ . The termini remain fixed, i.e.  $\mathbf{r}'_1 = \mathbf{r}_1$  and  $\mathbf{r}'_N = \mathbf{r}_N$ . The iterative procedure will continue to progressively smooth the chain. The main idea is to prevent the chains from passing through each other. This is done by checking that the triangles defined by  $\{\mathbf{r}_{i-1}, \mathbf{r}_i, \mathbf{r}'_i\}$  and  $\{\mathbf{r}_i, \mathbf{r}'_i, \mathbf{r}_{i+1}\}$  do not intersect any line segments defined by  $\{\mathbf{r}_{j-1}, \mathbf{r}_j\}$  for  $j < i$  and  $\{\mathbf{r}_j, \mathbf{r}_{j+1}\}$  for  $j > i$ . In practice, most protein chains reduce to a straight line defined either by two termini or to an obvious knot in less than 50 iterations. However, there are cases that will take 500 or more iterations to converge. Figure 1 shows a typical example of a chain-smoothing procedure from the original structure of the chromophore-binding domain of bacterial phytochrome (1ZTU) to the final smoothed chain that can be easily identified to contain a figure-eight knot.

### Data set and pre-computed knots

To speed up the web server, we pre-computed all proteins in the PDB as of January 12, 2007, which consists of 41 013 proteins comprising 34 971 X-ray structures and 6042 NMR protein structures. The crystal structures of homologous protein chains (even those with identical sequences) as well as the solution structures of the same protein were checked for the presence of knots. The chains with breaks or discontinuities are visually checked for their relevance in knot formation. If the proteins have a missing gap so large that it is improper to simply connect the two ends of the missing fragment to complete the chain, the identified knots will be disregarded. All final smoothed chains that appear to form a knot, i.e. not a simple straight line, were visually examined to decide whether these knots are authentic knots, slipknots or artificial knots caused by large breaks in the chains. The knots in proteins are quite simple in that they can be visually identified, and no sophisticated analysis [such as the Jones polynomials or others (20)] is required. In summary, pKNOT provides information about all knotted proteins,

such as their protein classes, their knotted types and the cores and depths of the knotted regions. The core is the smallest region that will remain knotted when the residues are successively deleted from both ends (15), and the depth is the product of the number of residues that must be deleted from both ends in order to free the knot (15).

Users can also upload the protein structure coordinates in the PDB format and the pKNOT server will progressively smooth the chains on the fly and then present the final smoothed chain as well as the original chain in a JAVA-based 3D graphics viewer AstexViewer (21) for users to inspect.

### Input format

The web page of the pKNOT web server is shown in Figure 2. The users can either type in the PDB ID or upload a structural file in the PDB format. In the latter case, the default iteration number is set to 500 and the collision threshold, to 0.5 Å. The user can either ignore or preserve the breaks in the chain when smoothing the chain. The former option will close the breaks by using the shortest line segment connecting the breaks, while the latter option preserves the breaks in the chain and smoothes each individual segment, keeping the endpoints of each segment fixed. The default is set to ignore the breaks in the chain. The users can also choose from the pull-down menu the number of iterations to smooth the chain. The collision threshold is the distance threshold to determine whether a line segment will intersect the triangle during the smoothing procedures.

### Output format and visualization of chains and knots

Upon query, pKNOT will return a table of the CHAIN, LENGTH, KNOT TYPE and DISPLAY STRUCTURE (Figure 3). When clicking on the column of KNOT TYPE, the server will return a list of all the proteins of the given knot type. pKNOT also provides the molecular viewer AstexViewer so that the users can visualize and manipulate in real time the protein structure and the knot in the protein. Both the original structure and the knot are

pKNOT

pKNOT<sub>1.0</sub>: Protein kNOT Server

## QUERY PROTEINS IN PDB

Enter PDB ID:

1T00

SUBMIT

## QUERY PROTEINS ON THE LOCAL MACHINE

## OPTIONS

- Ignore breaks in the chain  
 Preserve breaks in the chain

## OPTIONS

Iteration number 500  
 Collision threshold 0.50

Residue (specify the starting and the ending residues): 0 - 0

Submit protein structure (in PDB format):

Choose File no file selected

SUBMIT

**Figure 2.** The web page of the pKNOT web server. The users can either type in the PDB ID or upload a structural file in the PDB format. When submitting the structural file, the user can choose to either ignore or preserve the breaks in the chain. The default iteration number is set to 500 and the collision threshold, to 0.5Å.

shown in the same graphics window and the user can toggle on and off one of them for easy inspection.

## RESULTS

The knotted proteins come from the following protein classes: (1) methyltransferase, (2) transcarbamylase, (3) carbonic anhydrase, (3) ketol-acid reductoisomerase, (4) ubiquitin hydrolase, (5) methionine adenosyl transferase, (6) the chromophore-binding domain of bacterial phytochrome and (7) the inner core shell component protein of bluetongue virus. In addition, we also identified two knotted NMR structures: 1POQ and 1J2O. However, it is not clear whether these knots are authentic or due to incorrect structural refinement, since only one knotted model is identified among all NMR models for each protein (model 7 in 1POQ and model 14 in 1J2O).

## The knot types in proteins

There are three types of knot (up to the mirror image) identified in the PDB: the trefoil knot, the figure-eight knot and the knot with five crossings(15,19).

*The trefoil knot.* The trefoil knot (also called the threefoil or overhand knot) is the simplest knot of all, which is characterized by three crossings. It is mathematically denoted as a  $3_1$  knot. The proteins with a trefoil

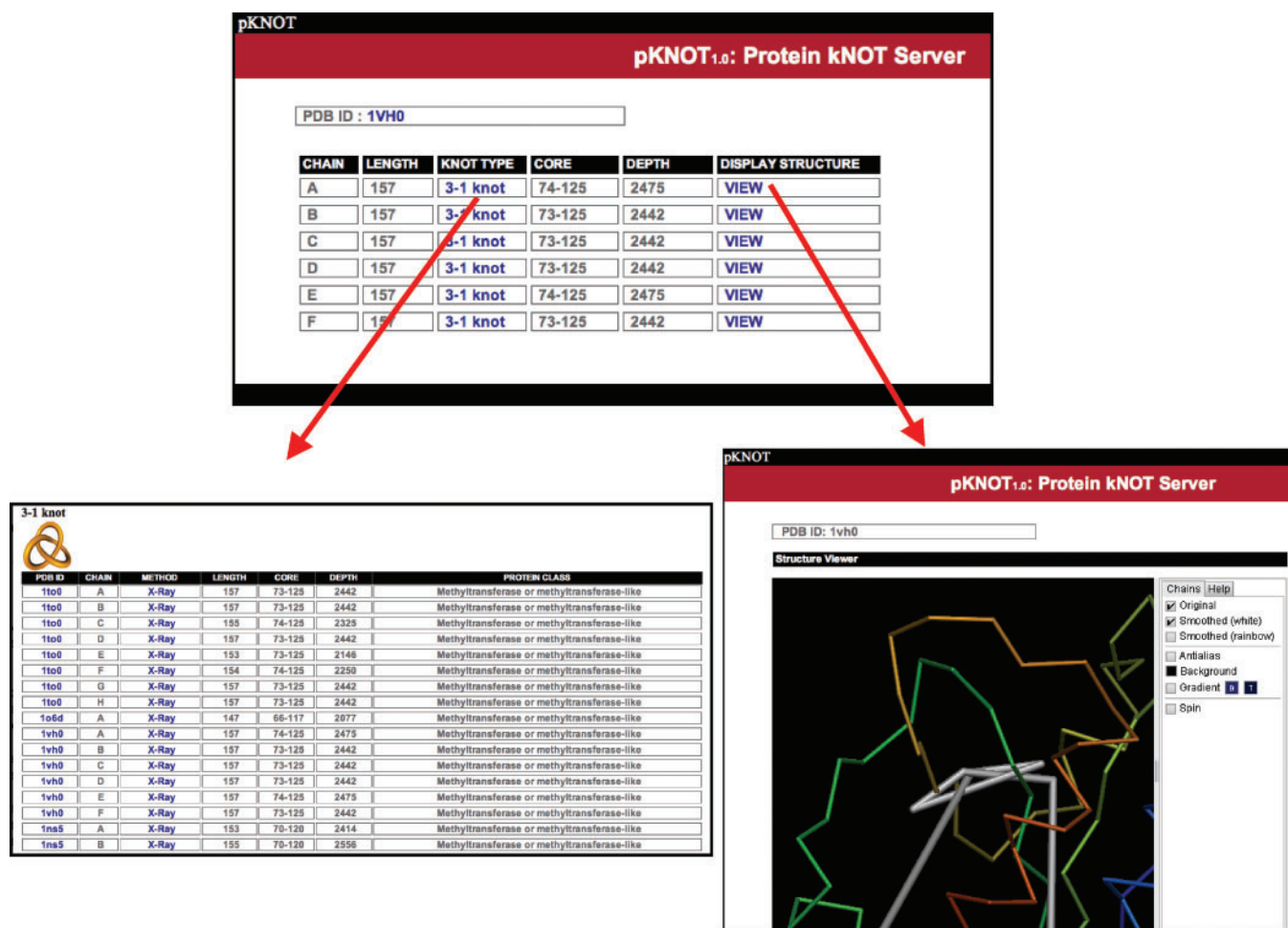
knot are (1) methyltransferase, (2) transcarbamylase, (3) methionine adenosyltransferase, (4) carbonic anhydrase and (5) YMPa superantigen (NMR).

*The figure-eight knot.* The figure-eight knot is characterized by four crossover points, alternately under and over. There is only one prime knot with four crossings and is denoted as the  $4_1$  knot. The proteins with a  $4_1$  knot are (1) the chromophore-binding domain of bacterial phytochrome, (2) the core protein of bluetongue virus, (3) ketol-acid reductoisomerase and (4) a LIM-ldbl-LID chimeric protein (NMR).

*The  $5_2$  knots.* There are two types of knot with five crossings: the  $5_1$  and  $5_2$  knots. Only the  $5_2$  knot has been identified in the protein structure and, as of writing, no proteins with six or more crossings have been identified in the PDB. The only protein family with a  $5_2$  knot is ubiquitin c-terminal hydrolase (1).

## Comparison with other work

It will be interesting to compare our results with those of the recent work by Lua and Grosberg (19). For example, they identified 19 knot proteins using the RANDOM method from the PDB-REPRDB data set (22) comprising 4716 representative protein. However, 5 of the identified 19 knotted proteins (1T0H:B, 1GKU:B, 1U2Z:C, 1M72:B



**Figure 3.** Upon query, the pKNOT server will return a table of the CHAIN, LENGTH, KNOT TYPE, CORE, DEPTH and DISPLAY STRUCTURE (upper center). When clicking on the column of KNOT TYPE, the server will return a list of all the proteins in the database of the pKNOT server (lower left). PKNOT also provides a JAVA-based 3D molecular viewer AstexViewer and the users can visualize and manipulate in real time the protein structure and the knot in the protein (lower right).

and 1X14:C) are questionable, since all of them have very large gaps in their structures due to missing residues. These knots arise either from the artificial virtual bonds that are used to connect the gaps or from the nonstandard PDB format. For example, 1T0H:B(23) has missing residues 414–424. A knot will form only if a virtual bond of length 32 Å connects the structural gap; 1U2Z:C has missing residues 570–573 and 575. The total distance of the structural gaps is around 52 Å. If these chain breaks were connected by virtual bonds, there will be a 4<sub>1</sub> knot. However, we notice that there is a chain in the complex (i.e. 1U2Z:A), which has identical sequence with 1U2Z:C and does not have a knot even if the structural gaps are connected by virtual bonds.

## CONCLUSION

Here we have presented the first web server to detect knots in proteins. With an increasing number of proteins with knots deposited in PDB, we believe that the pKNOT

web server will be useful to both biologists in general and structural biologists in particular.

## ACKNOWLEDGEMENTS

This work is supported in part by the National Science Council, the National Research Program of Genomic Medicine, and the MoE ATU plan. We are grateful to both the hardware and software supports of the Structural Bioinformatics Core Facility at National Chiao Tung University. We are also grateful to Dr. W. R. Taylor for providing the knot-detecting codes. Funding to pay the Open Access publication charge was provided by the MoE ATU plan.

*Conflict of interest statement.* None declared.

## REFERENCES

- Ahn, H.J., Eom, S.J., Yoon, H.J., Lee, B.I., Cho, H. and Suh, S.W. (2003) Crystal structure of class I acetoxyhydroxy acid



- isomeroeductase from *Pseudomonas aeruginosa*. *J. Mol. Biol.*, **328**, 505–515.
2. Ahn,H.J., Kim,H.W., Yoon,H.J., Lee,B.I., Suh,S.W. and Yang,J.K. (2003) Crystal structure of tRNA(m1G37)methyltransferase: insights into tRNA recognition. *EMBO J.*, **22**, 2593–2603.
  3. Elkins,P.A., Watts,J.M., Zalacain,M., van Thiel,A., Vitazka,P.R., Redlak,M., Andraos-Selim,C., Rastinejad,F. and Holmes,W.M. (2003) Insights into catalysis by a knotted TrmD tRNA methyltransferase. *J. Mol. Biol.*, **333**, 931–949.
  4. Lim,K., Zhang,H., Tempczyk,A., Krajewski,W., Bonander,N., Toedt,J., Howard,A., Eisenstein,E. and Herzberg,O. (2003) Structure of the YibK methyltransferase from *Haemophilus influenzae* (HI0766): a cofactor bound at a site formed by a knot. *Proteins*, **51**, 56–67.
  5. Mosbacher,T.G., Bechthold,A. and Schulz,G.E. (2005) Structure and function of the antibiotic resistance-mediating methyltransferase AviRb from *Streptomyces viridochromogenes*. *J. Mol. Biol.*, **345**, 535–545.
  6. Wagner,J.R., Brunzelle,J.S., Forest,K.T. and Vierstra,R.D. (2005) A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature*, **438**, 325–331.
  7. Liu,J., Wang,W., Shin,D.H., Yokota,H., Kim,R. and Kim,S.H. (2003) Crystal structure of tRNA (m1G37) methyltransferase from *Aquifex aeolicus* at 2.6 Å resolution: a novel methyltransferase fold. *Proteins*, **53**, 326–328.
  8. Forouhar,F., Shen,J., Xiao,R., Acton,T.B., Montelione,G.T. and Tong,L. (2003) Functional assignment based on structural analysis: crystal structure of the yggJ protein (HI0303) of *Haemophilus influenzae* reveals an RNA methyltransferase with a deep trefoil knot. *Proteins*, **53**, 329–332.
  9. Badger,J., Sauder,J.M., Adams,J.M., Antonysamy,S., Bain,K., Bergseid,M.G., Buchanan,S.G., Buchanan,M.D., Batiyenko,Y. *et al.* (2005) Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins*, **60**, 787–796.
  10. Shi,D., Morizono,H., Yu,X., Roth,L., Caldovic,L., Allewell,N.M., Malamy,M.H. and Tuchman,M. (2005) Crystal structure of N-acetylornithine transcarbamylase from *Xanthomonas campestris*: a novel enzyme in a new arginine biosynthetic pathway found in several eubacteria. *J. Biol. Chem.*, **280**, 14366–14369.
  11. Tyagi,R., Duquerroy,S., Navaza,J., Guddat,L.W. and Duggleby,R.G. (2005) The crystal structure of a bacterial class II ketol-acid reductoisomerase: domain conservation and evolution. *Protein Sci.*, **14**, 3089–3100.
  12. Misaghi,S., Galaray,P.J., Meester,W.J., Ovaa,H., Ploegh,H.L. and Gaudet,R. (2005) Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate. *J. Biol. Chem.*, **280**, 1512–1520.
  13. Das,C., Hoang,Q.Q., Kreinbring,C.A., Luchansky,S.J., Meray,R.K., Ray,S.S., Lansbury,P.T., Ringe,D. and Petsko,G.A. (2006) Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1. *Proc. Natl Acad. Sci. USA*, **103**, 4675–4680.
  14. Komoto,J., Yamada,T., Takata,Y., Markham,G.D. and Takusagawa,F. (2004) Crystal structure of the S-adenosylmethionine synthetase ternary complex: a novel catalytic mechanism of S-adenosylmethionine synthesis from ATP and Met. *Biochemistry.*, **43**, 1821–1831.
  15. Taylor,W.R. (2000) A deeply knotted protein structure and how it might fold. *Nature*, **406**, 916–919.
  16. Mallam,A.L. and Jackson,S.E. (2005) Folding studies on a knotted protein. *J. Mol. Biol.*, **346**, 1409–1421.
  17. Taylor,W.R., Xiao,B., Gamblin,S.J. and Lin,K. (2003) A knot or not a knot? SETting the record 'straight' on proteins. *Comput. Biol. Chem.*, **27**, 11–15.
  18. Jacobs,S.A., Harp,J.M., Devarakonda,S., Kim,Y., Rastinejad,F. and Khorasanizadeh,S. (2002) The active site of the SET domain is constructed on a knot. *Nat. Struct. Biol.*, **9**, 833–838.
  19. Lua,R.C. and Grosberg,A.Y. (2006) Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Comput. Biol.*, **2**, e45.
  20. Lickorish,W.B.R. (1997) *An Introduction to Knot Theory*. Springer, New York.
  21. Hartshorn,M.J. (2002) AstexViewer: a visualisation aid for structure-based drug design. *J. Comput. Aided Mol. Des.*, **16**, 871–881.
  22. Noguchi,T. and Akiyama,Y. (2003) PDB-REPRDB: a database of representative protein chains from the Protein Data Bank (PDB) in 2003. *Nucleic Acids Res.*, **31**, 492–493.
  23. Van Petegem,F., Clark,K.A., Chatelain,F.C. and Minor,D.L. Jr. (2004) Structure of a complex between a voltage-gated calcium channel beta-subunit and an alpha-subunit domain. *Nature*, **429**, 671–675.