Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 331 (2005) 56–60
BBRC
www.elsevier.com/locate/ybbrc

Modeling the tertiary structure of human cathepsin-E

Kuo-Chen Chou

Gordon Life Science Institute, 13784 Torrey Del Mar, San Diego, CA 92130, USA

Received 9 March 2005
Available online 30 March 2005

Abstract

Cathepsin-E is an endolysosomal aspartic proteinase and is predominantly expressed in immune system cells. Deficiency of cathepsin-E is associated with the development of atopic dermatitis, a pruritic inflammatory skin disease, which has put us to face a high selectivity challenge in the development of drugs for the therapy of Alzheimer's disease or breast cancer. This is because BACE1 (also known as β-secretase) and cathepsin-D, both belonging to the family of aspartic proteinases, might interact with the same compound as cathepsin-E does. BACE1 is a putative prime therapeutic target for the treatment of Alzheimer's disease, and cathepsin-D a potential target for breast cancer. Accordingly, in the course of finding drugs against Alzheimer's disease or breast cancer by inhibiting BACE1 or cathepsin-D, the desired drugs should selectively inhibit only BACE1 or cathepsin-D, but definitely not cathepsin-E. To realize this, it is indispensable to find out the structural difference of the three enzymes. Since the crystal structures of BACE1 and cathepsin-D are already known, the lack of three-dimensional structure of cathepsin-E has become the bottleneck in this regard. In view of this, the three-dimensional structure of cathepsin-E has been developed. Although the overall structures of the three enzymes are quite similar to each other, some subtle difference around their active sites that distinguishes cathepsin-E from cathepsin-D and BACE1 has been revealed through an analysis of hydrogen bond network and microenvironment. The computed three-dimensional structure of cathepsin-E and the relevant findings might provide useful insights for designing inhibitors with the desired selectivity.

Keywords: Alzheimer's disease; Atopic dermatitis; Breast cancer; Aspartic protease; BACE1; Cathepsin-D; Microenvironment; Hydrogen bond network; Selectivity

Cathepsin-E, like cathepsin-D and β-secretase (also known as BACE1), belongs to the category of aspartic proteases. However, owing to their different biological functions, they are targeted for completely different diseases with quite different strategies.

Cathepsin-D enhances anchorage-independent cell proliferation and subsequently facilitates tumorigenesis and metastasis of breast cancer cells, indicating the essential role of cathepsin-D in breast cancer and suggesting that cathepsin-D might be a target for finding drugs against breast cancer [1]. In view of this, we should inhibit cathepsin-D in order to block breast cancer.

BACE1 is predominantly expressed in the brain. It has been observed [2] that BACE knockout mice totally abolished the production of Aβ, the hallmark of Alzheimer's disease [3,4]. Accordingly, BACE1 is an attractive drug target for Alzheimer's disease therapy [5–7].

Cathepsin-E is an intracellular aspartic proteinase expressed predominantly in immune cells and skin. It has been reported [8] that cathepsin-E deficiency might be responsible for the induction of atopic dermatitis, a pruritic inflammatory skin disease. Thus, for the therapy of atopic dermatitis, instead of inhibiting it, we should activate cathepsin-E and make it function properly.

Therefore, in the process of finding drugs against breast cancer or Alzheimer's disease by inhibiting cathepsin-D or BACE1, the potential drugs must not inhibit cathepsin-E; otherwise, the disease of atopic

E-mail address: kchou@san.rr.com

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.03.123

dermatitis might be induced. In other words, the designed drugs must have very high selectivity. To realize this, we need to know the 3D (dimensional) structures of the protease domain for the three enzymes. For cathepsin-D and BACE1, the 3D structures of their protease domains have already been determined [6,9]. But the 3D structure of cathepsin-E is still unknown. The present study was initiated in an attempt to model the 3D structure of cathepsin-E, in the hope to provide useful information for conducting structure-based drug design to find compounds that only inhibit BACE1 or cathepsin-D, but not cathepsin-E.

Method

The sequence of human cathepsin-E was taken from [10]. The sequence consists of 396 residues, of which residues 1–17 belong to the signal peptide, residues 18–53 the pre-domain, and residues 54–396 the mature chain. In this study, the homologous technique was used for 3D structure prediction. The problem is what structure-known protein should be used as a template. There are two choices for the template: one is the crystal structure of BASE1, i.e., 1fkn.pdb [6]; and the other is the crystal structure of cathepsin-D, i.e., 1lya.pdb [9].

To properly select a template for cathepsin-E, it is instructive to perform a sequence similarity analysis of cathepsin-E with BACE1 and cathepsin-D, respectively. The sequence of human BACE1 was taken from Vassar et al. [11], and the sequence of human cathepsin-D was taken from Faust et al. [12]. The sequence of BACE1 consists of 501 residues, of which residues 1–21 is signal peptide, residues 22–57 the pro-domain, residues 58–446 the mature chain, residues 447–478 the transmembrane domain, and residues 479–501 the cytoplasmic domain. The sequence of cathepsin-D consists of 412 residues, of which residues 1–18 is signal peptide, residues 19–64 pro-domain, and residues 65–412 the mature chain. Since the current study is focused on the protease domain, the sequence similarity analysis should be performed for the mature chains only. It was found by using GAP of the GCG package [13] that the similarity and identity between cathepsin-E and BACE1 are 42.23% and 29.76%, while the corresponding rates between cathepsin-E and cathepsin-D are 62.76% and 52.55%, respectively. The results obtained by the sequence similarity analysis indicate that cathepsin-E has much higher sequence similarity and identity to cathepsin-D than to BACE1. Besides, it was found in the sequence alignments that between cathepsin-E and cathepsin-D there are six match positions with Cys; but between cathepsin-E and BACE1, only two such match positions. The folding of a protein is constrained by disulfide bond interaction. Therefore, it is more appropriate to choose the crystal structure of cathepsin-D [9] as a template to model the 3D structure for cathepsin-E.

The sequence alignment among cathepsin-E (cate), cathepsin-D (catd), and BACE1 (1fkn) performed by PILEUP in the GCG package [13] is given in Fig. 1, where codes in white characters with black background represent the conserved residues at the active site for the aspartyl protease family. Their sequence locations are residues 96–98 and residues 281–283 for cathepsin-E, residues 97–99 and residues 295–297 for cathepsin-D, and residues 93–95 and residues 289–291 for BACE1. The mature chain (protease domain) for cathepsin-E starts at residue 54 and ends at residue 396, that for cathepsin-D starts at residue 65 and ends at residue 412, and that for BACE1 starts at residue 58 and ends at residue 446. The codes connected by a solid line indicate the Cys pair involved in forming disulfide bond as observed in cathepsin-D. The nine residue segment shown in a box is the missing segment that has been automatically cleaved during the proteolytical process and hence is not included in the crystal structure.

Based on the sequence alignment (Fig. 1) and the X-ray co-ordinates of cathepsin-D (1lya.pdb) [9], the 3D structure of cathepsin-E was derived by the "segment matching" or "co-ordinate reconstruction" approach [14–18]. The modeling approach, which was based on the finding that most hexapeptide segments of protein structures could be clustered into only 100 structurally different classes, consisted of the following steps: (1) The target chain was first broken into many short sequence segments. (2) The database, which contains more than 5200 high-resolution crystal protein structures, was searched for matching the segments according to the sequence alignment of Fig. 1 and the "guiding" positions of the template protein chain (1lya.pdb). (3) These segment co-ordinates were fitted into the growing target structure under the monitor to avoid any van der Waals overlap until all atomic co-ordinates of the target structure were obtained. (4) The process was repeated 10 times and an average model was generated, followed by energy minimization to create the final 3D structure. The segment matching approach was previously used to model the structure of the protease domain of caspase-8, at a time before the crystal co-ordinates of caspase-3 were released [19]. To deal with such a situation, the 3D structure of the catalytic domain of caspase-3 was first derived by using the crystal structure of caspase-1 as a template; the caspase-3 structure thus obtained was subsequently used as a template to further model the protease domain of caspase-8. After the crystal co-ordinates of caspase-3 protease domain were finally released and the crystal structure of the caspase-8 protease domain was determined [20], it was found that the RMSD (root-mean-square-deviation) for all the backbone atoms of the caspase-3 protease domain between the crystal and computed structures was 2.7 Å, while the corresponding RMSD was 3.1 Å for caspase-8, and only 1.2 Å for its core structure, indicating that the predicted structures of caspase-3 and -8 were quite close to the corresponding crystal structures. Later on, this method was successively applied to model the caspase recruitment domains (CARDs) of Apaf-1, Ced-4, and Ced-3 by using the NMR structure of the RAIDD CARD [21] as a template, and to model the Cdk5–Nck5a* complex [22] as well as the protease domain of caspase-9 [23]. Two years after the computed Cdk5–Nck5a* complex structure was published [22], the corresponding crystal structure was determined [24]. It was quite encouraging to find that the predicted Cdk5 and the crystal Cdk5 are almost the same. Besides, according to the report for the crystal structure, upon the binding of Cdk5 and Nck5a* (or p25), the buried surface area was 3400 $Å^2$ [24], which was very close to 3461 $Å^2$ derived from the computed structure [22,25]. Also, stimulated by the computed Cdk5–Nck5a*–ATP structure, the molecular truncation experiments were conducted and it has been reported that the experimental results "confirm and extend specific aspects of the original predicted computer model" [25,26]. The segment match approach was also used to develop the 3D structures of extracellular domains for the subtypes 1, 2, 3, and 5 of GABA-A receptors [27], and the results thus obtained can be successfully used to clarify the long-standing ambiguity regarding the subunit arrangement (clockwise or counterclockwise) in the heteropentamers, providing useful insights for understanding the molecular action mechanism of the receptors.

Results and discussion

The predicted 3D structure thus obtained for the human cathepsin-E is illustrated in Fig. 2, where the mature chain is shown with ribbon drawing and colored green. The residues involved in forming the active site (DTG 96–98, DTG 281–283) and disulfide bonds are shown with the ball-and-stick drawing. The former is

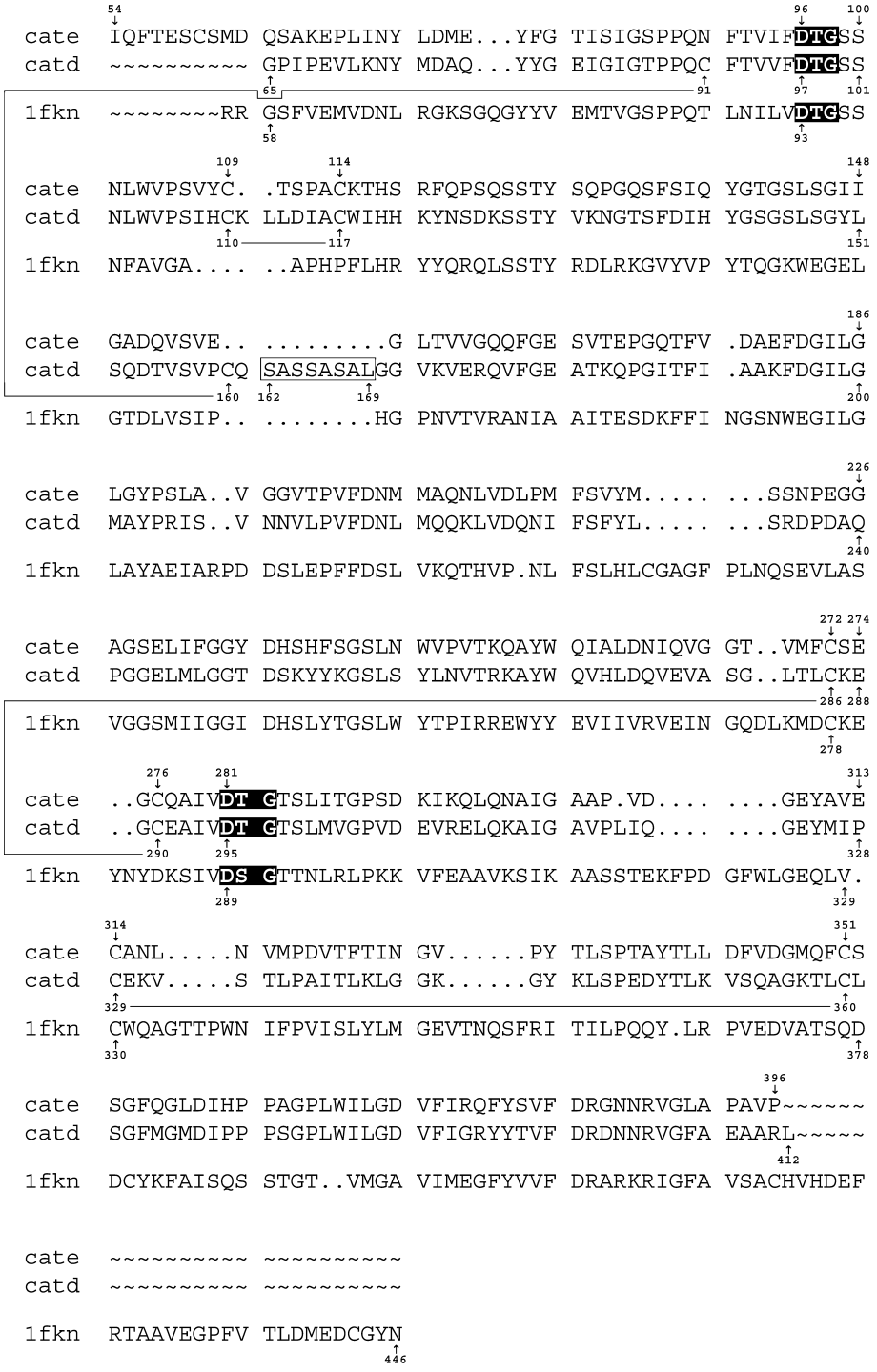

Fig. 1. Sequence alignment of cathepsin-E (cate), cathepsin-D (catd), and BACE1 (1fkn). The alignment was performed using PILEUP of GCG package [13]. Only mature chains of the three proteins were included during the operation. The codes in white characters with black background represent the conserved residues at the active site for the aspartyl protease family. Their sequence locations are residues 96–98 and residues 281–283 for cathepsin-E, residues 97–99, and residues 295–297 for cathepsin-D, and residues 93–95 and residues 289–291 for BACE1. The two codes connected by a solid line represent the Cys pair involved in forming a disulfide bond as observed in the crystal structure of cathepsin-D determined by Baldwin et al. [9].

colored red and the latter colored yellow. As shown in Fig. 3, the structure contains two pairs of disulfide bonds, i.e., Cys^{272}–Cys^{276} and Cys^{314}–Cys^{351}. Although Cys^{109} and Cys^{114} are quite close to each other, the distance between their two S^{γ} atoms is 3.53 Å, unable to form a good disulfide bond. The 3D model can provide us with a detailed look at the structural difference of cathepsin-E with BACE1 and cathepsin-D.

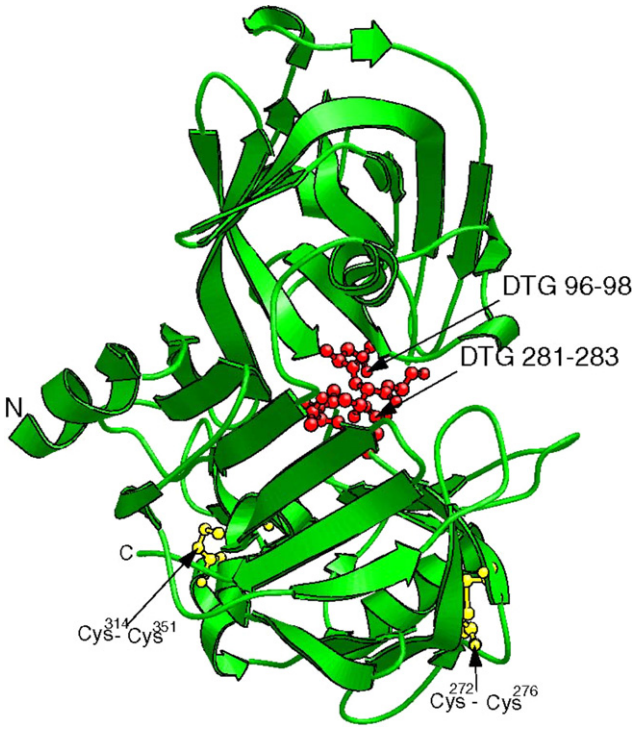

Fig. 2. The computed structure of cathepsin-E based on the sequence alignment of Fig. 1 and the crystal structure of cathepsin-D (1lya.pdb). The main chain is shown by ribbon drawing. The residues at the DTG/DTG active site and those involved in forming disulfide bonds are shown with ball-and-stick drawing; the former is colored red and the latter yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

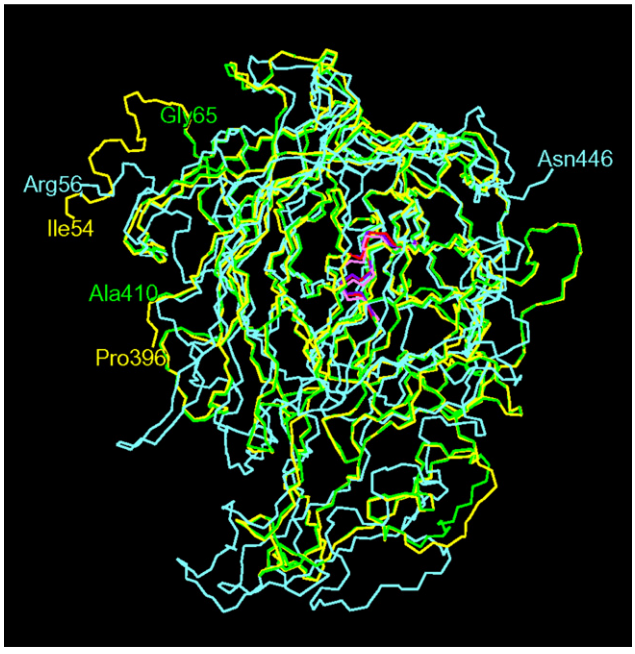

Fig. 3. A backbone superposition of BACE1 (light blue), cathepsin-D (green), and cathepsin-E (yellow). The corresponding DTG/DSG segments are colored pink, purple, and red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

A backbone superposition of the crystal structure of BACE1 [6], crystal structure of cathepsin-D [9], and the predicted structure of cathepsin-E is given in Fig. 3, where BACE1 is colored blue with active site DTG (residues 93–95) and DSG (residues 289–291) in pink, cathepsin-D colored green with active site DTG (residues 97–99) and DTG (residues 295–297) in purple, and cathepsin-E colored yellow with DTG (residues 96–98) and DTG (residues 281–283) in red. Also, the N-terminal and C-terminal residues for each of the three structures are marked by the same color as that of the corresponding chain. It can be seen from the superimposed structure that the backbone conformation of cathepsin-E is much closer to cathepsin-D than to BACE1. Nevertheless, the backbone conformations of the DTG/DSG active sites for the three enzymes are almost identical to each other.

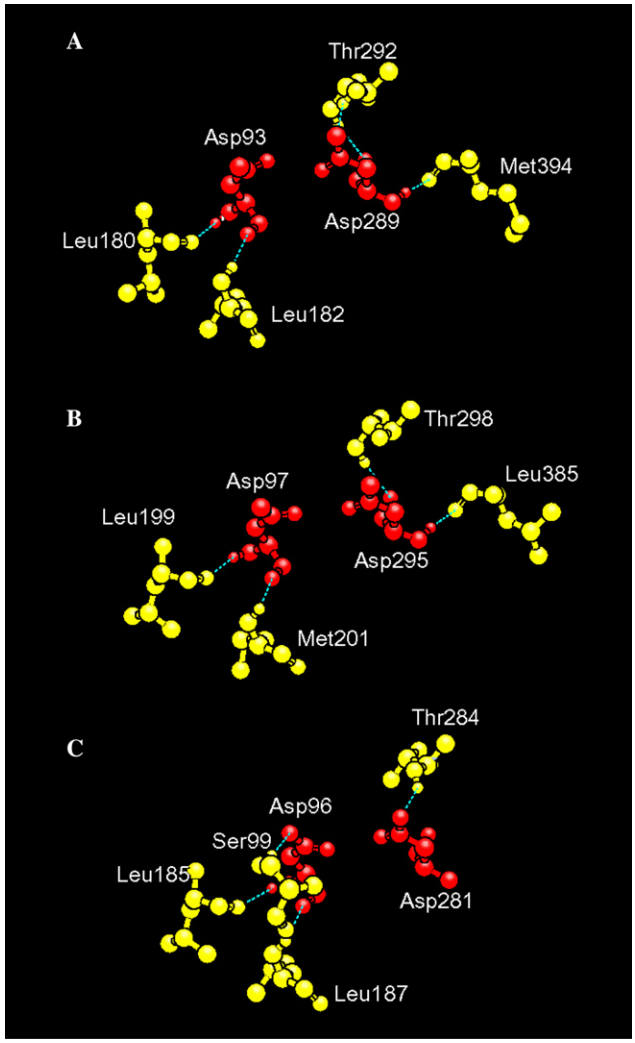

Fig. 4. An illustration to show the microenvironment of the two active site Asp residues for (A) BACE1, (B) cathepsin-D, and (C) cathepsin-E. Only heavy atoms, as well as those hydrogen atoms involved in forming hydrogen bonds, are shown.

An analysis about the microenvironments around the two active site Asp residues for the three enzymes is given in Fig. 4, where panel (A) is for BACE1, panel (B) for cathepsin-D, and panel (C) for cathepsin-E. As we can see from the figure, there are five hydrogen bonds between the two active site Asp residues and the surrounding residues for BACE1, but only four such hydrogen bonds for cathepsin-D and E, respectively. The hydrogen bond network has a common feature, i.e., one active site Asp residue always forms two hydrogen bonds with Lus/Met residues; while the other Asp forms one or two hydrogen bonds with Thr residue. A special feature that distinguishes cathepsin-E from BACE1 and cathepsin-D is that Asp96 in cathepsin-E forms an additional hydrogen bond with Ser99, and Asp281 does not form hydrogen bond with Leu/Met residue as the corresponding Asp residues in BACE1 and cathepsin-D do. This kind of subtle difference might be of use for designing a highly selective inhibitor.

Conclusion

The overall 3D structures of cathepsin-D, BACE1, and cathepsin-E are very similar to each other. However, there are differences in the microenvironments around their active sites. These differences may provide useful insight for designing highly selective drugs for therapy without causing abnormal effects.

References

[1] M. Glondu, E. Liaudet-Coopman, D. Derocq, N. Platet, H. Rochefort, M. Garcia, Oncogene 21 (2002) 5127–5134.
[2] S.L. Roberds, J. Anderson, G. Basi, Hum. Mol. Genet. 10 (2001) 1317–1324.
[3] J. Hardy, D. Allsop, Trends Pharmacol. Sci. 12 (1991) 383–388.
[4] D.B. Carter, K.C. Chou, Neurobiol. Aging 19 (1998) 37–40.
[5] J.T. Huse, D.S. Pijak, G.J. Leslie, J. Biol. Chem. 275 (2000) 33729–33737.
[6] L. Hong, G. Koelsch, X. Lin, S. Wu, S. Terzyan, A.K. Ghosh, X.C. Zhang, J. Tang, Science 290 (2000) 150–153.
[7] K.C. Chou, W.J. Howe, Biochem. Biophys. Res. Commun. 292 (2002) 702–708.
[8] T. Tsukuba, K. Okamoto, Y. Okamoto, M. Yanagawa, K. Kohmura, Y. Yasuda, H. Uchi, T. Nakahara, M. Furue, K. Nakayama, T. Kadowaki, K. Yamamoto, K.I. Nakayama, J. Biochem. 134 (2003) 893–902.
[9] E.T. Baldwin, T.N. Bhat, S. Gulnik, M.V. Hosur, R. Sowder, R.E. Cachau, J. Collins, A.M. Silva, J.W. Erickson, Proc. Natl. Acad. Sci. USA 90 (1993) 6796–6800.
[10] J. Hill, D.S. Montgomery, J. Kay, FEBS Lett. 326 (1993) 101–104.
[11] R. Vassar, B.D. Bennett, S. Babu-Khan, S. Kahn, E.A. Mendiaz, P. Denis, D.B. Teplow, S. Ross, P. Amarante, R. Loeloff, Y. Luo, S. Fisher, J. Fuller, S. Edenson, J. Lile, M.A. Jarosinski, A.L. Biere, E. Curran, T. Burgess, J.-C. Louis, F. Collins, J. Treanor, G. Rogers, M. Citron, Science 286 (1999) 735–741.
[12] P.L. Faust, S. Kornfeld, J.M. Chirgwin, Proc. Natl. Acad. Sci. USA 82 (1985) 4910–4914.
[13] J. Devereux, Genetic Computer Group (GCG), Madison, Wisconsin, 1994.
[14] R. Unger, D. Harel, S. Wherland, J.L. Sussman, Proteins: Struct. Funct. Bioinform. 5 (1989) 355–373.
[15] M. Claessens, E.V. Cutsem, I. Lasters, S. Wodak, Protein Eng. 4 (1989) 335–345.
[16] T.A. Jones, S. Thirup, EMBO J. 5 (1986) 819–822.
[17] T.L. Blundell, B.L. Sibanda, M.J.E. Sternberg, J.M. Thornton, Nature (London) 326 (1987) 347–352.
[18] K.C. Chou, G. Nemethy, M. Pottle, H.A. Scheraga, J. Mol. Biol. 205 (1989) 241–249.
[19] K.C. Chou, D. Jones, R.L. Heinrikson, FEBS Lett. 419 (1997) 49–54.
[20] W. Watt, K.A. Koeplinger, A.M. Mildner, R.L. Heinrikson, A.G. Tomasselli, K.D. Watenpaugh, Structure 7 (1999) 1135–1143.
[21] J.J. Chou, H. Matsuo, H. Duan, G. Wagner, Cell 94 (1998) 171–180.
[22] K.C. Chou, K.D. Watenpaugh, R.L. Heinrikson, Biochem. Biophys. Res. Commun. 259 (1999) 420–428.
[23] K.C. Chou, A.G. Tomasselli, R.L. Heinrikson, FEBS Lett. 470 (2000) 249–256.
[24] C. Tarricone, R. Dhavan, J. Peng, L.B. Areces, L.H. Tsai, A. Musacchio, Mol. Cell 8 (2001) 657–669.
[25] K.C. Chou, Curr. Med. Chem. 11 (2004) 2105–2134.
[26] J. Zhang, C.H. Luan, K.C. Chou, G.V.W. Johnson, Proteins: Struct. Funct. Genet. 48 (2002) 447–453.
[27] K.C. Chou, Biochem. Biophys. Res. Commun. 316 (2004) 636–642.