A structural tree for proteins containing 3β-corners

A.V. Efimov*

Institute of Protein Research, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

Received 5 February 1997; revised version received 3 March 1997

Abstract A structural tree for β-proteins with predominantly orthogonal β-sheet packing has been constructed. The 3β-corner, a structural motif that recurs in proteins of this class, is taken as a root structure of the tree. The 3β-corner can be represented as a triple-stranded β-sheet folded on to itself so that its two β-β-hairpins are packed approximately orthogonally in different layers and the central strand bends by ~90° in a right-handed direction when passing from one layer to the other. The larger protein structures are obtained by stepwise addition of β-strands to the root 3β-corner taking into account a restricted set of rules inferred from known principles of protein structure. The protein structures that can be obtained in this way are grouped into one structural class and those found in branches of the structural tree into subclasses.

Key words: Protein classification; Protein structure comparison; Stepwise folding; Structural motif; Structural similarity

1. Introduction

Modelling of folding pathways of globular proteins may be of particular value in understanding the principles that govern the polypeptide chain folding. One approach is based on stepwise addition of α-helices and/or β-strands to the corresponding initiating complex [1,2] or structural motif with a unique overall fold [3–5] taking into account a restricted set of simple rules. A folding scheme that includes all the intermediate and final structures connected by lines showing possible folding pathways may be represented as a structural tree with the corresponding root structural motif [4]. The number of possible overall folds that can be obtained from one structural motif is limited since the rules drastically reduce the number of allowed pathways of growth of intermediate structures. Thus, the structural tree is a good tool for searching of all possible protein folds as well as for structure comparison and protein classification. Levels of structural similarity between different proteins and domains can easily be observed by visual inspection of the trees. Proteins and domains found within a structural tree can be grouped into one structural class and those found within branches of the tree into subclasses. In addition, novel structural features in proteins may be observed and some novel rules may be inferred from analysis of structural relatedness of proteins.

Structural trees for five superfamilies, β-proteins containing abcd units, α-proteins containing α-α-corners, two-layer α/β-proteins containing abCd units, three-layer α/β-proteins containing five-segment and seven-segment α/β-motifs, have been constructed previously [4]. This paper describes a structural tree for proteins containing 3β-corners and implications for protein classification.

2. Construction and analysis of the structural tree

The 3β-corner is a structural motif that can be represented as a triple-stranded β-sheet folded on to itself so that its two β-β-hairpins are packed approximately orthogonally in different layers and the central strand bends by ~90° in a right-handed direction when passing from one layer to the other [5]. All the 3β-corners observed in proteins, when viewed from their concave surfaces, can be considered as formed by Z-like β-sheets. In other words, the first and second strands form a right-turned β-β-hairpin and the second and third strands a left-turned β-β-hairpin when the 3β-corner is viewed from the concave surface. The 3β-corners are widespread in both homologous and non-homologous proteins and domains and some small proteins are merely composed of 3β-corners and short irregular 'tails' (see Fig. 1). All this suggests that the 3β-corner can adopt its unique structure per se and can be a core around which the remainder of the molecule or domain is folded. So the 3β-corner is taken as the starting structure in protein modelling and the root structure in constructing the structural tree.

The structural tree has been constructed taking into account a restricted set of general rules that have been derived from analysis of the structural features observed in globular proteins:

(1) Overall folds of protein molecules and intermediate structures are taken into account and details of the structures are ignored.

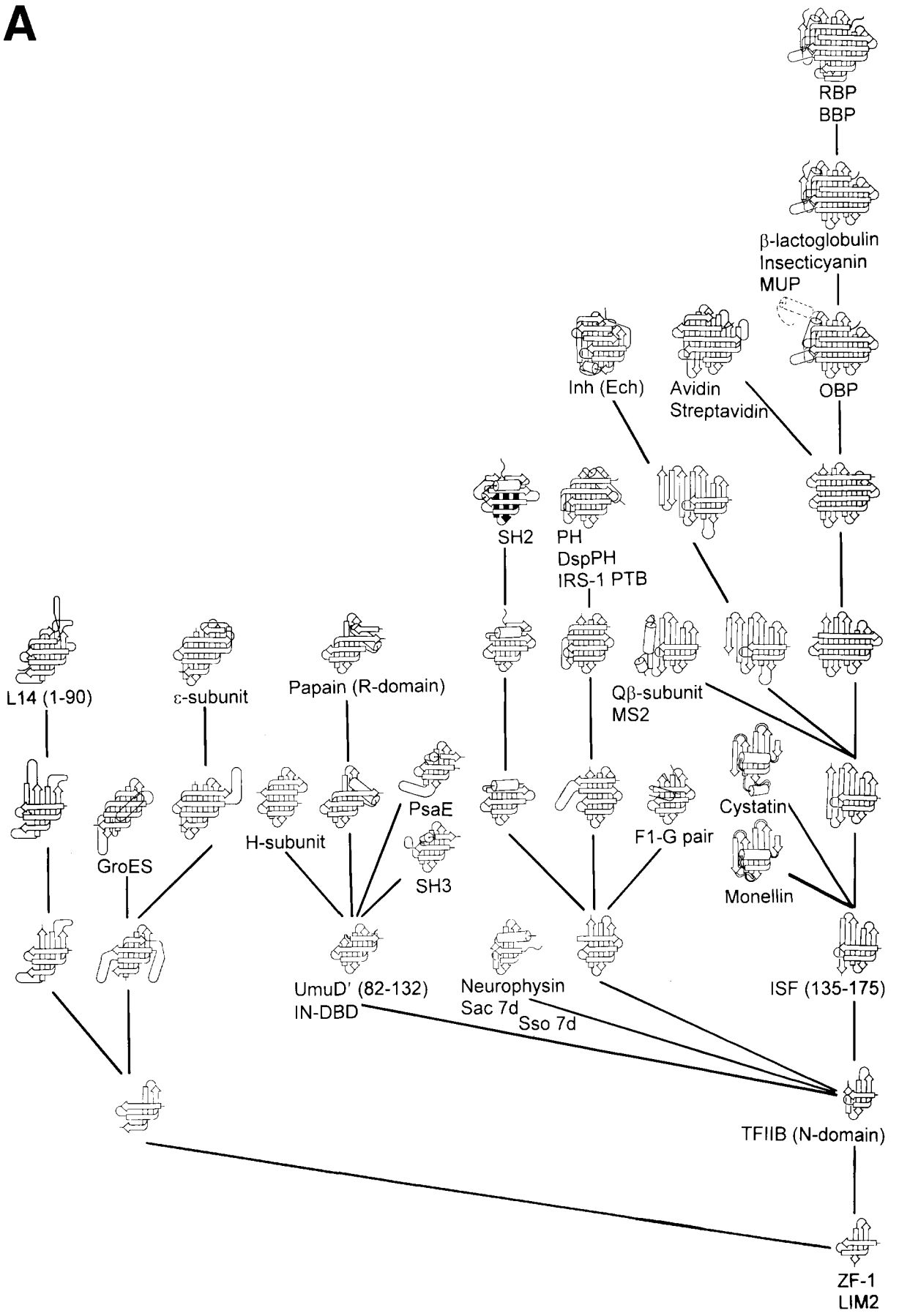
(2) The larger protein and intermediate structures are obtained by stepwise addition of β-strands and α-helices to a growing structure so that a structure obtained at the preceding step is maintained (it can be slightly modified). At each step, the β-strand or α-helix nearest to the growing structure along the chain is the first to be attached to the growing structure [3–5].

(3) The obtained structures should be compact; α-helices and β-strands should be packed in accordance with the rules that govern their close packing (see, e.g. [6,7]).

(4) α-Helices and β-strands cannot be packed into one layer because of dehydration of the free NH and CO groups of the β-strands [6]. Close packings of α-helices and β-strands into one layer results in some main-chain NH and CO groups of β-strands being unable to have partners to form intramolecular hydrogen bonds and becoming inaccessible for water molecules and, consequently, dehydrated. From a thermodynamic point of view, this is very unfavorable. Thus, a β-strand should be packed into the β-layer and an α-helix into the α-helical layer or into a hydrophobic concavity (it can be formed, for example, by a strongly twisted and coiled β-sheet

*Corresponding author. Fax: (7) (095) 924-0493.
E-mail: efimov@ipr.serpukhov.su

PII S0014-5793(97)00296-2


A
RBP
BBP
β-lactoglobulin
Insecticyanin
MUP
Inh (Ech)
Avidin
Streptavidin
OBP
SH2
PH
DspPH
IRS-1 PTB
L14 (1-90)
ε-subunit
Papain (R-domain)
Qβ-subunit
MS2
PsaE
H-subunit
F1-G pair
Cystatin
GroES
SH3
Monellin
UmuD' (82-132)
IN-DBD
Neurophysin
Sac 7d
Sso 7d
ISF (135-175)
TFIIB (N-domain)
ZF-1
LIM2

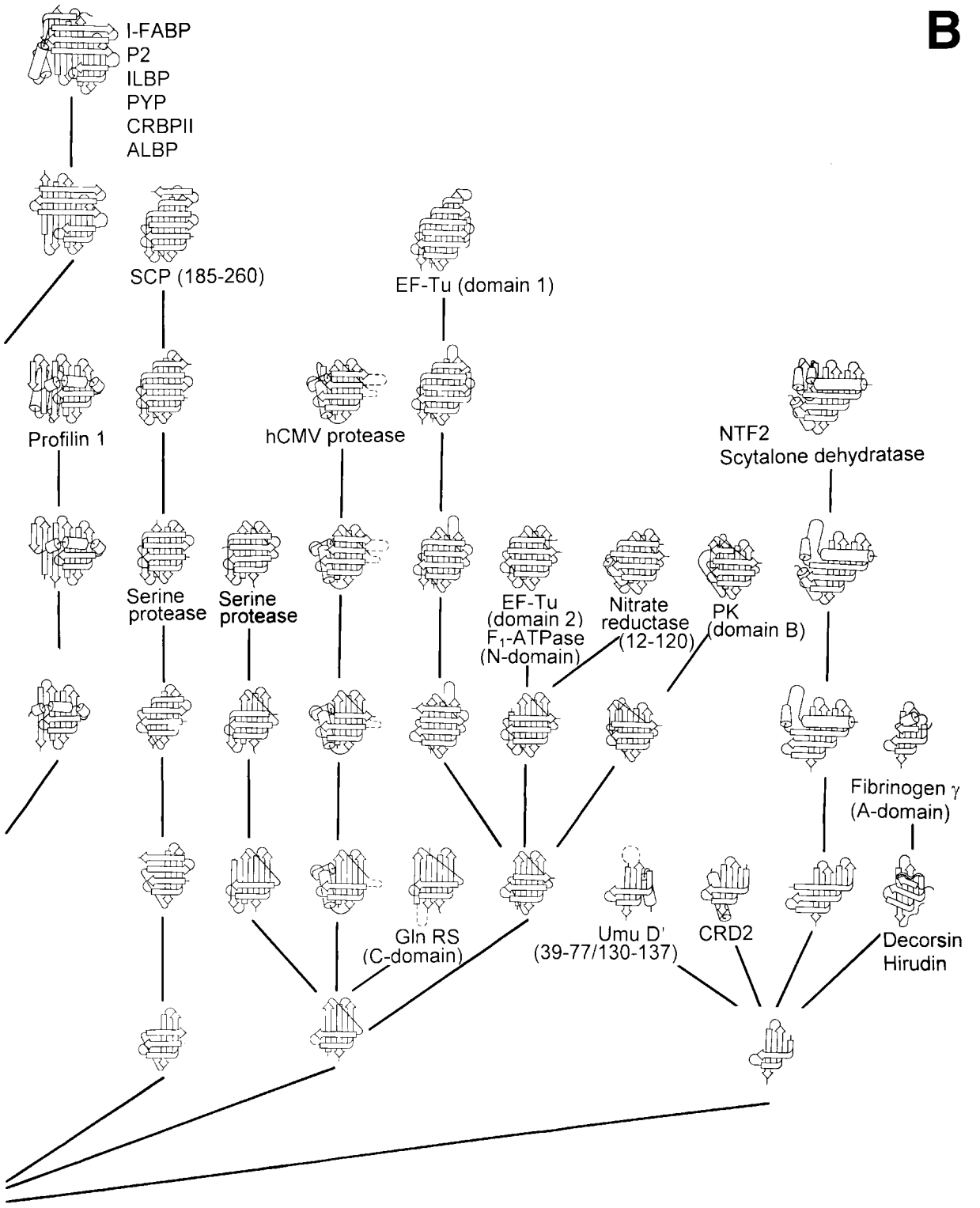

B
I-FABP
P2
ILBP
PYP
CRBPII
ALBP
SCP (185-260)
EF-Tu (domain 1)
Profilin 1
hCMV protease
NTF2
Scytalone dehydratase
Serine protease
Serine protease
EF-Tu (domain 2) F1-ATPase (N-domain)
Nitrate reductase (12-120)
PK (domain B)
Fibrinogen γ (A-domain)
Gln RS (C-domain)
Umu D' (39-77/130-137)
CRD2
Decorsin
Hirudin

Fig. 1. A structural tree for proteins and domains containing 3β-corners. β-Strands are shown as arrows directed from N- to C-ends and α-helices as cylinders. The β-strands of the near β-sheets are oriented horizontally and those of the far β-sheets vertically. Long loops are simplified and drawn by dashed lines. Segments that can be present or not present in a structure are also shown by dashed lines. Thick lines show possible pathways of stepwise growth of the root 3β-corner and intermediate structures.The structural infirmation is taken from the following papers: L14, ribosomal protein L14 [10]; Gro ES monomer [11]; ε-subunit of ATP synthase [12]; H-subunit of the photosynthetic reaction center [13]; Umu D′ protein [14]; IN-DBD, the DNA binding domain of HIV-1 integrase [15]; Psa E, a photosystem I protein [16]; SH-3 domain [17]; papain [18]; neurophysin [19]; Sac 7d, the DNA binding protein Sac 7d [20]; Sso 7d, the DNA binding protein Sso 7d [21]; SH2 domain [22]; PH, the pleckstrin homology domain [23]; Dsp PH, the PH domain from β-spectrin [24]; IRS-1 PTB, the PTB domain of insulin receptor substrate-1 [25]; F1-G pair [26]; cystatin [27]; monellin, single-chain monellin [28]; Qβ-subunit, a subunit of phage Qβ capsid [29]; MS2, coat protein of bacteriophage MS2 [30]; Ihn (Ech), *Erwinia chrysanthemi* inhibitor [31]; ZF-1, synthetic zinc-finger peptide 1 [32]; LIM 2 domain [33]; TFIIB, transcription initiation factor TFIIB [34]; ISF, iron-sulfur fragment of cytochrome bc1 [35]; OBP, bovine odorant binding protein [36]; avidin [37]; streptavidin [38]; β-lactoglobulin [39]; insecticyanin [40]; MUP, major urinary protein [41]; RBP, retinol binding protein [42]; BBP, bilin binding protein [43]; I-FABP, intestinal fatty acid binding protein [44]; P2, P2 myelin protein [45]; ILBP, ileal lipid binding protein [46]; PYP, photoactive yellow protein [47]; CRBP II, cellular retinol binding protein II [48]; ALBP, adipocyte lipid binding protein [49]; profilin I [50]; serine protease, a trypsin-like serine protease (see, e.g. [51]); SCP, Sindbis virus core protein [52]; hCMV protease [53]; Gln RS, glutaminyl-tRNA synthetase [54]; EF-Tu, elongation factor EF-Tu [55]; F1-ATPase [56]; PK, pyruvate kinase [57]; nitrate reductase [58]; NTF 2, nuclear transport factor 2 [59]; scytalone dehydratase [60]; hirudin [61]; CRD2, cysteine-rich domain 2 [62]; decorsin [63]; domain A (143–191) of the C-terminal fragment of the γ-chain of fibrinogen [64] also has the decorsin-like fold but contains a short α-helix 152–160 between the first and second β-strands.
←

or between two β-sheets that splay apart) of a growing structure.

(5) Crossing of connections [8] and formation of knots [9] are prohibited.

(6) All the 3β-corners (not only root 3β-corners) and other structural motifs should have corresponding handedness and overall folds [3–5]. β-Strands that covalently link the two β-sheets at close corners and bend through 90° when passing from one β-sheet to the other should form so-called right-handed bends [7]. Left-handed bends have longer loops and could occur only at corners where the two layers of the β-sandwich splay apart [7].

Fig. 1 shows a structural tree constructed in accordance with these rules. All the structures are oriented in a similar way so that root 3β-corners are localized in their bottom right corners and the β-strands of the near β-sheets are directed horizontally and those of the far β-sheets vertically. There are two β-layers packed approximately orthogonally in the root 3β-corner. So each subsequent β-strand can be packed into one or the other β-layer of a growing structure. As can be seen, addition of a β-strand to the root 3β-corner at the first step can be done in different ways and results in formation of the structures shown in the bottom row of the tree. The next row represents the structures obtained by addition of two β-strands or one β-strand and one α-helix to the root 3β-corner, etc. All the pathways of stepwise growth of the root 3β-corner that lead to known protein structures are shown with thick lines.

As can be seen, the structural tree has several branches. Within one branch, structures having a higher position in the tree include the structures located lower. Proteins and domains of different branches have a common fold located in the branching point. The higher a branching point is located in the tree, the higher the level of structural similarity between proteins and domains of the corresponding branches.

Proteins and domains found within this structural tree can be grouped into one structural class or superfamily, proteins and domains containing 3β-corners. Proteins and domains found within branches of the tree can be considered subclasses or subfamilies (see also [3,4] for other classes and subclasses). This classification is different from those suggested by other authors [65–67] in some aspects. First of all, amino acid sequences and functions of proteins are not taken into account in this classification. It is primarily based on similarity of overall folds and modelled pathways of stepwise growth of the motifs. As seen, this approach permits the structural classification of both homologous and non-homologous proteins and offers a stimulating perspective regarding their folding pathways.

Analysis of all data presented above has led us to a hypothesis that the 3β-corner can fold independently of the remaining parts of the molecules and can act as a nucleus in protein folding, and the pathways of its stepwise growth can be considered possible folding pathways of proteins and domains containing it.

Acknowledgements: This work was supported in part by the Russian Foundation for Basic Research (RFFI Grant #95-04-11851a).

References

[1] O.B. Ptitsyn, A.V. Finkelstein, P. Falk (Bendzko), FEBS Lett 101 (1979) 1–5.
[2] O.B. Ptitsyn, A.V. Finkelstein, Q Rev Biophys 13 (1980) 339–386.
[3] A.V. Efimov, Mol Biol (in Russian) 16 (1982) 799–806.
[4] Efimov AV. Proteins 1997 (in press).
[5] A.V. Efimov, FEBS Lett 298 (1992) 261–265.
[6] A.V. Efimov, Dokl Akad Nauk SSSR 235 (1977) 699–702.
[7] C. Chothia, J. Janin, Biochemistry 21 (1982) 3955–3965.
[8] V.I. Lim, A.L. Mazanov, A.V. Efimov, Mol Biol (in Russian) 12 (1978) 206–213.
[9] J.S. Richardson, Nature 268 (1977) 495–500.
[10] C. Davies, S.W. White, V. Ramakrishnan, Structure 4 (1996) 55–66.
[11] J.F. Hunt, A.J. Weaver, S.J. Landry, L. Gierasch, J. Deisenhofer, Nature 379 (1996) 37–45.
[12] S. Wilkens, F.W. Dahlquist, L.P. McIntosh, L.W. Donaldson, R.A. Capaldi, Nature Struct Biol 2 (1995) 961–967.
[13] J. Deisenhofer, O. Epp, K. Miki, R. Huber, H. Michel, Nature 318 (1985) 618–624.
[14] T.S. Peat, E.G. Frank, J.P. McDonald, A.S. Levine, R. Woodgate, W.A. Hendrickson, Nature 380 (1996) 727–730.
[15] A.P.A.M. Eijkelenboom, R.A.P. Lutzke, R. Boelens, R.H.A. Plasterk, R. Kaptein, K. Hård, Nature Struct Biol 2 (1995) 807–810.
[16] C.J. Falzone, Y.-H. Kao, J. Zhao, D.A. Bryant, J.T.J. Lecomte, Biochemistry 33 (1994) 6052–6062.
[17] A. Masacchio, M. Noble, R. Pauptit, R. Wierenga, M. Saraste, Nature 359 (1992) 851–855.
[18] J. Drenth, J.N. Jansonius, R. Koekoek, B.G. Wolthers, Adv Protein Chem 25 (1971) 79–115.
[19] L. Chen, J.P. Rose, E. Breslow, D. Yang, W.-R. Chang, W.F.

Furey Jr., M. Sax, B.-C. Wang, Proc Natl Acad Sci USA 88 (1991) 4240–4244.
[20] S.P. Edmondson, L. Qiu, J.W. Shriver, Biochemistry 34 (1995) 13289–13304.
[21] H. Baumann, S. Knapp, T. Lundbäck, R. Ladenstein, T. Härd, Nature Struct Biol 1 (1994) 808–819.
[22] M.J. Eck, S.E. Shoelson, S.C. Harrison, Nature 362 (1993) 87–91.
[23] D. Timm, K. Salim, I. Gout, L. Guruprasad, M. Waterfield, T. Blundell, Nature Struct Biol 1 (1994) 782–788.
[24] P. Zhang, S. Talluri, H. Deng, D. Branton, G. Wagner, Structure 3 (1995) 1185–1195.
[25] M.-M. Zhou, B. Huang, E.T. Olejniczak, R.P. Meadows, S.B. Shuker, M. Miyazaki, T. Trüb, S.E. Shoelson, S.W. Fesik, Nature Struct Biol 3 (1996) 388–393.
[26] B.O. Smith, A.K. Downing, P.C. Driscoll, T.J. Dudgeon, I.D. Campbell, Structure 3 (1995) 823–833.
[27] W. Bode, R. Engh, D. Musil, U. Thiele, R. Huber, A. Karshikov, J. Brzin, J. Kos, V. Turk, EMBO J 7 (1988) 2593–2599.
[28] J.P. Somoza, F. Jiang, L. Tong, C.-H. Kang, J.M. Cho, S.-H. Kim, J Mol Biol 234 (1993) 390–404.
[29] R. Golmohammadi, K. Fridborg, M. Bundule, K. Valegård, L. Liljas, Structure 4 (1996) 543–554.
[30] K. Valegård, L. Liljas, K. Fridborg, T. Unge, Nature 345 (1990) 36–41.
[31] U. Baumann, M. Bauer, S. Létoffé, P. Delepelaire, C. Wandersman, J Mol Biol 248 (1995) 653–661.
[32] A. Hammarström, K.D. Berndt, R. Sillard, K. Adermann, G. Otting, Biochemistry 35 (1996) 12723–12732.
[33] G.C. Pérez-Alvarado, C. Miles, J.W. Michelsen, H.A. Louis, D.R. Winge, M.C. Beckerle, M.F. Summers, Nature Struct Biol 1 (1994) 388–398.
[34] W. Zhu, Q. Zeng, C.M. Colangelo, L.M. Lewis, M.F. Summers, R.A. Scott, Nature Struct Biol 3 (1996) 122–124.
[35] S. Iwata, M. Saynovits, T.A. Link, H. Michel, Structure 4 (1996) 567–579.
[36] M. Tegoni, R. Ramoni, E. Bignetti, S. Spinelli, C. Cambillau, Nature Struct Biol 3 (1996) 863–867.
[37] L. Pugliese, M. Malcovati, A. Coda, M. Bolognesi, J Mol Biol 235 (1994) 42–46.
[38] W.A. Hendrickson, A. Pähler, J.L. Smith, Y. Satow, E.A. Merrit, R.P. Phizackerley, Proc Natl Acad Sci USA 86 (1989) 2190–2194.
[39] M.Z. Papiz, L. Sawyer, E.E. Elipoulos, A.C.T. North, J.B.C. Findlay, R. Sivaprasadarao, T.A. Jones, M.E. Newcomer, P.J. Kraulis, Nature 324 (1986) 383–385.
[40] H.M. Holden, W.R. Rypniewski, J.H. Law, I. Rayment, EMBO J 6 (1987) 1565–1570.
[41] Z. Böcskei, C.R. Groom, D.R. Flower, C.E. Wright, S.E.V. Philips, A. Gavaggioni, J.B.C. Findlay, A.C.T. North, Nature 360 (1992) 186–188.
[42] G. Zanotti, S. Ottonello, R. Berni, H.L. Monaco, J Mol Biol 230 (1993) 613–624.
[43] R. Huber, M. Schneider, I. Mayr, R. Müller, R. Deutzmann, F. Suter, H. Zuber, H. Falk, H. Kayser, J Mol Biol 198 (1987) 499–513.
[44] J.C. Sacchettini, J.I. Gordon, L.J. Banaszak, J Mol Biol 208 (1989) 327–339.
[45] T.A. Jones, T. Bergfors, J. Sedzik, T. Unge, EMBO J 7 (1988) 1597–1604.
[46] C. Lücke, F. Zhang, H. Rüterjans, J.A. Hamilton, J.C. Sacchettini, Structure 4 (1996) 785–800.
[47] D.E. McRee, J.A. Tainer, T.E. Meyer, J. Van Beeumen, M.A. Cusanovich, E.D. Getzoff, Proc Natl Acad Sci USA 86 (1989) 6533–6537.
[48] N.S. Winter, J.M. Bratt, L.J. Banaszak, J Mol Biol 230 (1993) 1247–1259.
[49] Z. Xu, D.A. Bernlohr, L.J. Banaszak, Biochemistry 31 (1992) 3484–3492.
[50] A.A. Fedorov, K.A. Magnus, M.H. Graupe, E.E. Lattman, T.D. Pollard, S.C. Almo, Proc Natl Acad Sci USA 91 (1994) 8636–8640.
[51] G.H. Cohen, E.W. Silverton, D.R. Davies, J Mol Biol 148 (1981) 449–479.
[52] S. Lee, K.E. Owen, H.-K. Choi, H. Lee, G. Lu, G. Wengler, D.T. Brown, M.G. Rossmann, R.J. Kuhn, Structure 4 (1996) 531–541.
[53] L. Tong, C. Qian, M.-J. Massariol, P.R. Bonneau, M.G. Cordingley, L. Lagacé, Nature 383 (1996) 272–275.
[54] M.A. Rould, J.J. Perona, D. Söll, T.A. Steitz, Science 246 (1989) 1135–1142.
[55] M. Kjeldgaard, J. Nyborg, J Mol Biol 223 (1992) 721–742.
[56] M.J. Van Raaij, J.P. Abrahams, A.G.W. Leslie, J.E. Walker, Proc Natl Acad Sci USA 93 (1996) 6913–6917.
[57] A. Mattevi, G. Valentini, M. Rizzi, M.L. Speranza, M. Bolognesi, A. Coda, Structure 3 (1995) 729–741.
[58] G. Lu, W.H. Campbell, G. Schneider, Y. Lindqvist, Structure 2 (1994) 809–821.
[59] T.L. Bullock, W.D. Clarkson, H.M. Kent, M. Stewart, J Mol Biol 260 (1996) 422–431.
[60] T. Lundqvist, J. Rice, C.N. Hodge, G.S. Basarab, J. Pierce, Y. Lindqvist, Structure 2 (1994) 937–944.
[61] T. Szyperski, P. Güntert, S.R. Stone, K. Wüthrich, J Mol Biol 228 (1992) 1193–1205.
[62] U. Hommel, M. Zurini, M. Luyten, Nature Struct Biol 1 (1994) 383–387.
[63] A.M. Krezel, G. Wagner, J. Seymour-Ulmer, R.A. Lazarus, Science 264 (1994) 1944–1947.
[64] V.C. Yee, K.P. Pratt, H.C.F. Côté, I.L. Trong, D.W. Chung, E.W. Davie, R.E. Stenkamp, D.C. Teller, Structure 5 (1997) 125–138.
[65] C.A. Orengo, T.P. Flores, W.R. Taylor, J.M. Thornton, Protein Eng 6 (1993) 485–500.
[66] L. Holm, C. Sander, Nucleic Acids Res 22 (1994) 3600–3609.
[67] A.G. Murzin, S.E. Brenner, T. Hubbard, C. Chothia, J Mol Biol 247 (1995) 536–540.