PROTEINS: Structure, Function, and Bioinformatics 65:280–284 (2006)

Crystal Structure of a Consensus-Designed Ankyrin Repeat Protein: Implications for Stability

H. Kaspar Binz, Andreas Kohl, Andreas Plückthun,* and Markus G. Grütter*
Biochemisches Institut, Universität Zürich, Zürich, Switzerland

ABSTRACT **Consensus-designed ankyrin repeat (AR) proteins are thermodynamically very stable. The structural analysis of the designed AR protein E3_5 revealed that this stability is due to a regular fold with highly conserved structural motifs and H-bonding networks. However, the designed AR protein E3_19 exhibits a significantly lower stability than E3_5 (9.6 vs. 14.8 kcal/mol), despite 88% sequence identity. To investigate the structural correlations of this stability difference between E3_5 and E3_19, we determined the crystal structure of E3_19 at 1.9 Å resolution. E3_19 as well has a regular AR domain fold with the characteristic H-bonding patterns. All structural features of the E3_5 and E3_19 molecules appear to be virtually identical ($RMSD_{C\alpha} \approx 0.7$ Å). However, clear differences are observed in the surface charge distribution of the two AR proteins. E3_19 features clusters of charged residues and more exposed hydrophobic residues than E3_5. The atomic coordinates of E3_19 have been deposited in the Protein Data Bank. PDB ID: 2BKG. Proteins 2006;65:280–284.**

Key words: ankyrin repeat; consensus sequence; protein stability; protein design; repeat protein; X-ray crystallography

INTRODUCTION

Repeat proteins such as ankyrin repeat (AR), leucine-rich repeat (LRR), or tetratricopeptide repeat (TPR) proteins are abundant specific binding molecules in nature.[1–8] They are composed of small structurally homologous units (repeats) that stack to form a repeat domain. To biophysically analyze these nonglobular proteins and to use repeat proteins as binding molecules, repeat proteins were recently subjected to protein engineering involving consensus design.[9–15]

We have constructed libraries of AR proteins of varying repeat numbers.[10] By sequence and structural analyses, we designed a 33 amino acid AR module with 27 defined consensus framework residues and 6 randomized potential interaction residues. Although 26 framework residues were fixed, one was allowed to be His, Tyr, or Asn. All randomized potential interaction residues were allowed to be any amino acid except Cys, Gly, or Pro. Varying numbers of this AR module were cloned between designed capping ARs, which are terminal ARs shielding the hydrophobic core of the AR domain. In this manner, combinatorial libraries were generated coding for AR proteins containing one N-terminal capping AR, two, three, or four AR modules and one C-terminal capping AR. To reflect this composition, the libraries were termed N2C, N3C, and N4C.

Our previous analyses showed that the designed AR proteins are well expressed, soluble, monomeric, and thermodynamically stable.[10,11] The crystal structure of the N3C protein E3_5 was determined at 2 Å resolution, and it displayed the typical AR protein fold. Moreover, the consensus design resulted in a highly regular AR protein structure. Interestingly, E3_5 is thermodynamically significantly more stable (ΔG = 14.8 kcal/mol, $T_m > 85$°C) than another N3C protein analyzed, E3_19 (ΔG = 9.6 kcal/mol, T_m = 66°C),[10,11] although the two proteins share 89% residue similarity and 88% identity (see Fig. 1). We now have determined the crystal structure of E3_19 at 1.9 Å resolution to find the structural reasons for these differences in stability. A detailed comparison of E3_19 with E3_5 is presented. In another paper by Yu et al.,[16] the comparison between E3_19 and E3_5 is further extended by molecular dynamics studies.

METHODS

Crystallization and Data Collection

E3_19 protein was expressed and purified as described[10,11] with an additional Superdex-75 (Pharmacia, Dübendorf, Switzerland) gel-filtration step in 10 mM Tris-HCl pH 7.6, 100 mM NaCl. E3_19 at a concentration of 18 mg/mL in this buffer crystallized in 1 week using the sitting drop-vapor diffusion method, in 96-well crystallization plates at 20°C. The drops contained 1 μL of protein and 1 μL of reservoir buffer (15% PEG 4000, 200 mM KBr, and 100 mM Tris-HCl, pH 8.0), with 100 μL of reservoir buffer in each well. For cryoprotection, crystals were dipped in paraffin oil, before flash freezing them at 90 K for data collection. X-ray diffraction was measured at the PX beamline at the Swiss Light Source (Villingen, Switzer-

Grant sponsor: the National Center of Competence in Research (NCCR) in Structural Biology.

*Correspondence to: Andreas Plückthun, Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. E-mail: plueckthun@bioc.unizh.ch, or Markus G. Grütter, Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. E-mail: gruetter@bioc.unizh.ch

Received 11 April 2005; Revised 20 October 2005; Accepted 1 November 2005

Published online 21 February 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20930

Fig. 1. Repeat-wise sequence alignment of E3_5 and E3_19. The two molecules share 88% sequence identity and 89% sequence similarity. The differences are exclusively found in randomized positions. Secondary structure elements are shown above the sequences, the repeat name is indicated on the right. Cylinders represent α-helices and two black bars represent β-turns. Note that the repeats were defined to end after the GADVNA motif to better illustrate the sequence differences in the β-turn region. The N-terminal MRGSHHHHHH sequence has been removed from the N-terminal capping repeat in this representation.

land) using a MarCCD detector (Mar USA Inc., Evanston, IL). A data set from a single crystal diffracting to 1.9 Å resolution was collected and processed with MOSFLM,[17] SCALA,[17] and TRUNCATE.[17] The crystal belonged to space group $P2_12_12_1$. A Matthews coefficient of $V_M = 2.11$ Å^3/Da (41.7% water content) was calculated using the molecular weight of 17,778 Da and assuming two molecules per asymmetric unit. Statistics on data collection are given in Table I.

Molecular Replacement, Model Building, and Refinement

The crystal structure was determined by molecular replacement using the program AMoRe[18] with the structure of the E3_5 (PDB ID: 1MJ0[11]) as a search model. A conventional AMoRe protocol (rotation and translation) was applied and yielded two solutions, where the two solutions were symmetry related copies. Model building was carried out using the program O.[19] The structure was refined in CNS,[20] followed by REFMAC,[21] resulting in a final model with an R-factor of 17.8% and an R_{free}-factor of 22.7%. Three hundred water molecules were inserted manually or using CNS. Refinement details are given in Table I. In the molecular replacement solution of E3_19, one of the two molecules in the asymmetric unit was frameshifted by one repeat, compared to the final model. The molecule was hence translated by one repeat towards the N-terminus.

Analysis and Bioinformatics

The model was evaluated using the programs PROCHECK[22] and WHAT IF.[23] HBPLUS[24] and LIGPLOT[25] were used for analysis of the H-bonding networks and the hydrophobic contacts; GRASP[26] was used for surface calculations and cavity search. Swiss PDB viewer[27] and MolMol[28] were used for root-mean-square deviation (RMSD) calculations and molecular model display.

TABLE I. Statistics for Data Collection and Refinement

Data collection	
Space group	$P2_12_12_1$
Cell dimensions, Å	$a = 49.579$, $b = 72.670$ $c = 83.325$ $\alpha = \beta = \gamma = 90.00°$
Resolution limits, Å	25–1.9
Observed reflections	total: 99379 unique: 24393
Completeness, %	99.9 (99.9)[a]
Redundancy	4.1
R_{sym} (% on I)	8.1 (34.4)[a]
Refinement	
Resolution range, Å	25–1.90
Rfactor/R_{free}, %	17.83/22.72
Ordered water molecules	300
rms deviation from ideal geometry:	
Bond lengths, Å	0.02
Bond angles, °	1.746
Average B factor, Å^2	17.027

[a]Number in parenthesis refers to the highest resolution shell.

RESULTS

Refinement and Crystal Structure Description

In the final model of E3_19, clear electron density extends from Ser12 to Leu165 in both molecules of the asymmetric unit, while the N-terminal RGS-His-tag (MRGSHHHHHHG) and the C-terminal Gln166 were not defined in either molecule. In the Ramachandran plot 91.9% of the residues were in the most favored region, 8.1% in the additionally allowed region, while no residue was in the generously allowed or disallowed region A PROCHECK[22] analysis further indicated that all main-chain and side-chain parameters were within the standard values. The two molecules in the asymmetric unit show the typical AR protein fold (Fig. 2) and are practically identical with a $RMSD_{C\alpha}$ of approximately 0.3 Å (Table II). An $RMSD_{C\alpha}$ comparison with E3_5[11] shows that E3_5 and E3_19 are almost identical ($RMSD_{C\alpha}$ approximately 0.7 Å; Table II). The largest differences are observed in the C-terminal AR (Fig. 2). Low $RMSD_{C\alpha}$ are also observed when comparing E3_19 to the AR protein off7 ($RMSD_{C\alpha} < 1.0$ Å),[9] which binds to maltose binding protein, or to GABPβ1 ($RMSD_{C\alpha} <$ 1.1 Å).[29] A repeat-by-repeat $RMSD_{C\alpha}$ analysis locates the largest (but still small) differences between E3_19 and E3_5 in the terminal repeats (Table II, Fig. 2), while the consensus-designed repeats are highly similar. The two molecules in the asymmetric unit are involved in a total of seven crystal contacts with a dominating contact between the A molecule and a symmetry copy of the B molecule. Similar to E3_5,[11] E3_19 features highly conserved inter- and intrarepeat H-bond patterns both in the β-turns and in the TPLH motif. The number of H-bonds of the two

TABLE II. Full Protein and Repeat $RMSD_{C\alpha}$ Comparison between E3_19 and E3_5

Fragments compared	N123C	N	1	2	3	C	123
E3_19a vs. E3_5	0.72	0.48	0.29	0.36	0.41	0.58	0.48
E3_19b vs. E3_5	0.63	0.50	0.23	0.28	0.24	0.56	0.43
E3_19a vs. E3_19b	0.32	0.32	0.11	0.15	0.14	0.01	0.19

molecules are virtually identical (29 per repeat in E3_19 vs. 30 in E3_5). The number of interrepeat H-bonds is slightly higher in E3_5 (4.75 on average vs. 4 in E3_19). Cavities in the molecules (1 with a volume of 20 $Å^3$) and hydrophobic contact areas between single repeats of E3_19 (1440 $Å^2$ each) are identical to E3_5. The randomized framework positions (residues Asn69, Tyr102, and Tyr135) are similarly oriented as in E3_5, and the two sequence differences in framework positions 102 and 135 (H102Y and H135Y) do not seem to influence the molecules. With 20.03 $Å^2$ (molecule A) and 18.66 $Å^2$ (molecule B), respectively, the overall *B*-factors of the two E3_19 molecules are very similar and slightly higher than the one of E3_5 (16.27 $Å^2$). The major differences in the *B*-factors compared to E3_5 are found in residues of the N-terminal ARs of E3_19, while the residues in the internal and C-terminal ARs exhibit comparable *B*-factors for all molecules.

Differences between E3_5 and E3_19

Although the overall structures of E3_19 and E3_5 are virtually identical, some details are noteworthy. The main differences between the two molecules are located in the composition of the randomized positions of the designed AR modules. Unlike E3_5, E3_19 shows a clustering of charges in the randomized potential interaction residues. In the β-turn region, where our AR proteins are negatively charged by design (four negative charges in the framework), E3_19 has three additional charges from the randomized residues and thus a total of 7 negative charges, which are more clustered than the six negative charges of E3_5 (two additional charges from randomized residues; Figs. 1 and 3). E3_19 also has an aspartate in the first α-helix of the fourth repeat, while E3_5 has no additional negative charges in the randomized helix positions. In the first α-helices of the first three repeats, E3_19 has four clustered positive charges (Fig. 3), while E3_5 has only two well-separated charges (N-terminal capping repeat and fourth repeat). E3_19 displays four hydrophobic residues, while E3_5 displays two (one semiburied). In E3_5, Asn158 of the C-terminal capping AR makes a double H-bond to Lys122 and Ala121 of the fourth repeat, while in E3_19, Asn156 takes over this role (bonding to Asp122; Fig 2). It is not inconceivable that this small alteration is a consequence of a repulsion between Asp122, Asp151, and Asp155 in E3_19. However, the shift in H-bonds and the simultaneous structural alteration in the C-terminal capping repeat might also be caused by crystal contacts.

DISCUSSION

The structural differences between E3_19 and E3_5 are relatively small and do not provide an explanation for differences between the thermodynamic stabilities of the two molecules. Sequence differences that do not alter the fold must therefore be the major reason for the large differences in thermodynamic stability. These sequence differences are exclusively due to the composition of the randomized positions. E3_19 has clusters of positive and negative charges and two more hydrophobic exposed residues than E3_5. In addition, the arrangement of the charges and the hydrophobic residues appears to be less favorable in E3_19 than in E3_5. It may also be that not only the arrangement of the charges in clusters but generally the location of the charges is unfavorable in E3_19, which would indicate the presence of hot spots in the randomized positions that are critical to the AR domain stability. It should be noted that there is a cluster of positive charges (mostly arginines) in adjacent helices in E3_19, while there are only few charges in these positions in E3_5. Also, the second to last repeat of E3_19 carries more negative charges (−5) than the corresponding repeat of E3_5 (−3), which could have an effect on the thermodynamic stability because the C-terminal AR also carries five negative charges. The paper by Yu et al.[16] will describe molecular dynamics studies and electrostatic calculations in which E3_19 is compared to E3_5 quantitatively.

Previously, we also studied two different N2C library AR proteins (named E2_5 and E2_17).[10,11] These two proteins are more similar (91% identity, 93% similarity) to each other than E3_5 is to E3_19, and the composition of their randomized positions is more balanced. At the same time their difference in stability is smaller (ΔT_m = 9°C; $\Delta\Delta G$ = 1.87 kcal/mol). This supports the notion that differences in thermodynamic stability are caused by the composition of the randomized residues, and it further indicates that the stability difference is particularly pronounced in the E3_5/E3_19 pair.

In summary, we have compared the crystal structures of the still highly stable AR protein E3_19 (ΔG = 9.6 kcal/mol) with the even more stable AR protein E3_5 (ΔG = 14.8 kcal/mol) to find the structural reasons for the differences in the stabilities. The randomized residues of these molecules do not cause obvious changes in the AR domain fold, but the resulting amino acid composition does influence the thermodynamic stability of the AR domain. Due to their identical fold, E3_19 and E3_5 could represent a suitable model system for molecular dynamics simulations to evaluate the contribution of single amino acids to the stability of the AR domain fold.

ACKNOWLEDGMENTS

We thank Dr. Christophe Briand, Andreas Schweizer, Daniel Frey, Christian Stirnimann, Heinz Gut, and Dr. Guido Capitani for support and helpful comments. We are further grateful to Dr. Patrik Forrer, Dr. Patrick Amstutz,

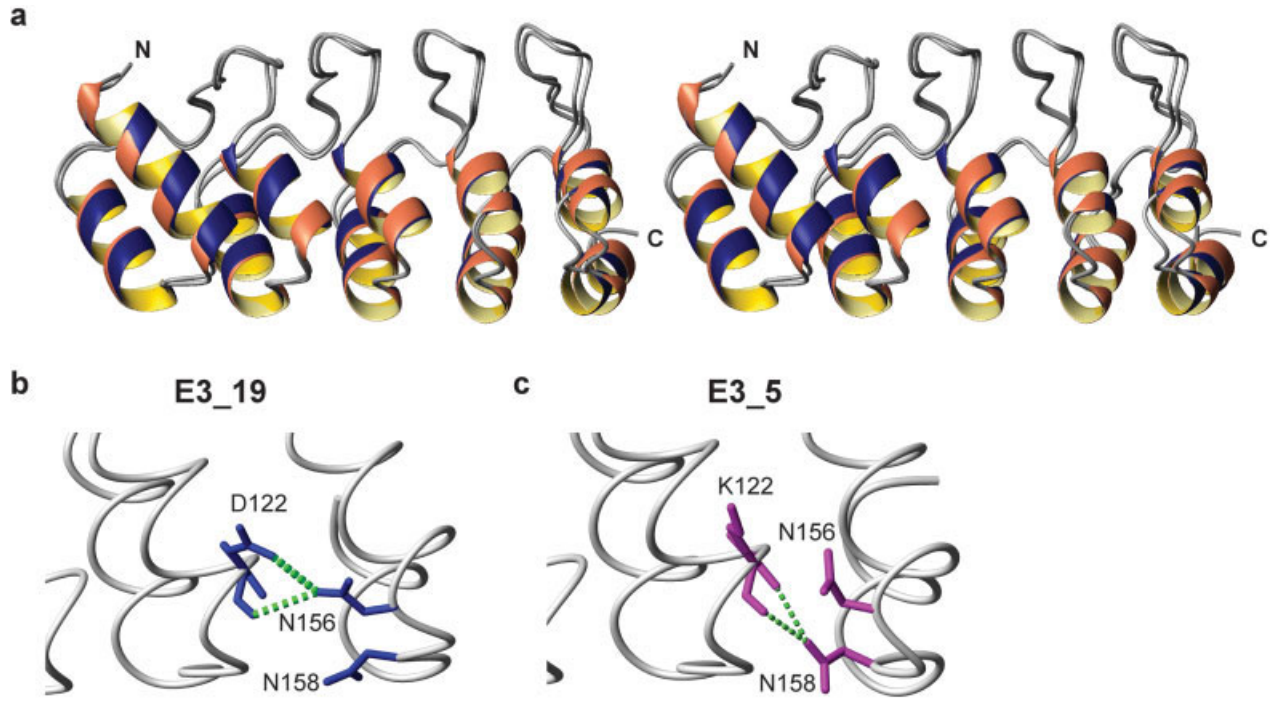

Fig. 2. Structural comparison of of E3_19 and E3_5. (**a**) Stereoview of the superposition of E3_19 (B molecule, brown) with E3_5 (PDB ID: 1MJ0, blue) in ribbon representations. The N- and C-termini are labeled for orientation. The two proteins are practically identical, with the biggest differences being observed in the C-terminal AR, where E3_5 appears to be more compact than E3_19. The two E3_19 molecules of the asymmetric unit are virtually identical (**b**) H-bonding in the C-terminal AR of E3_19. Asn156 makes H-bonds to Asp122 (residues in stick mode in blue). (**c**) H-bonding in the C-terminal AR of E3_5. Asn158 makes H-bonds to Lys122 and Ala121 (residues in stick mode in magenta). This figure was prepared using MolMol.[28]

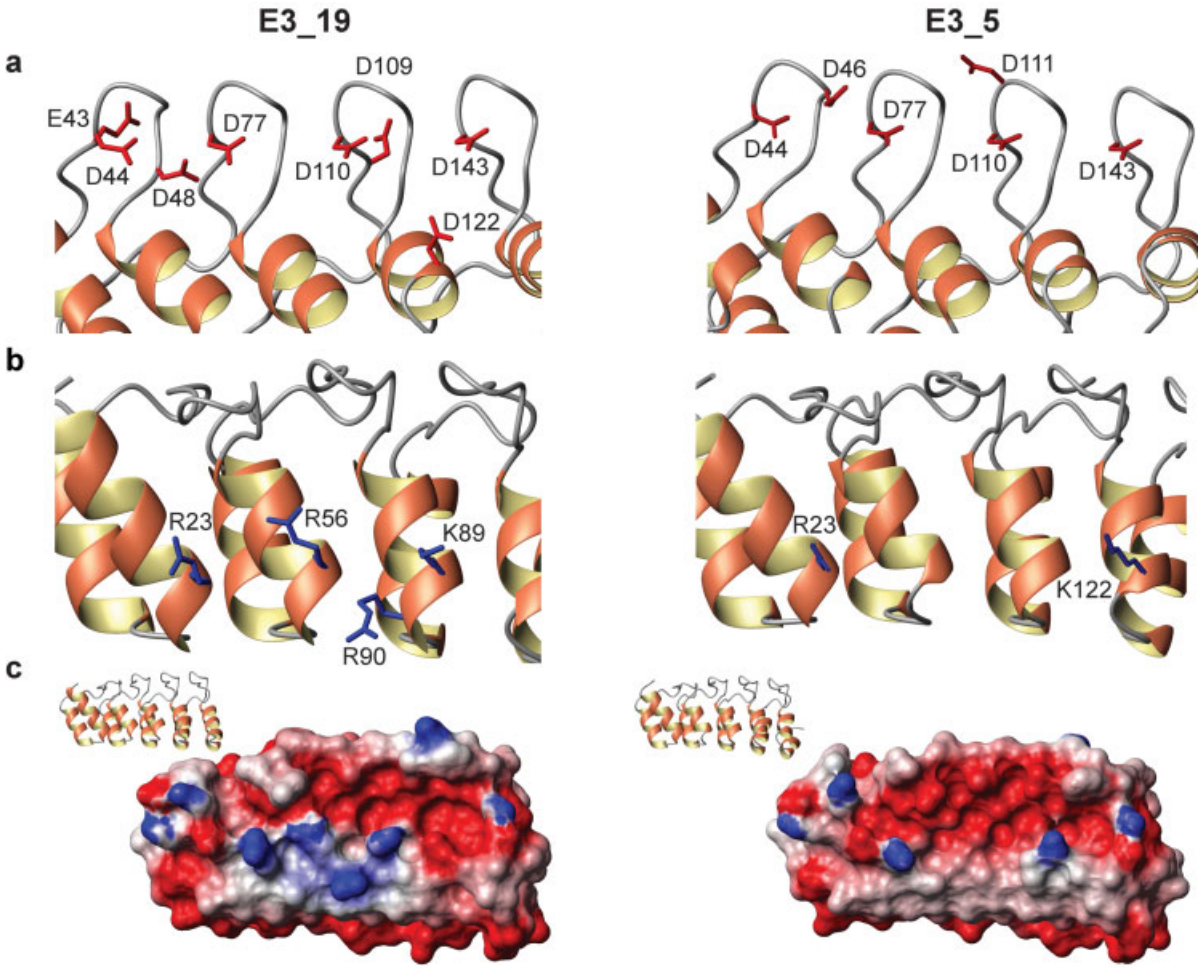

Fig. 3. Charge distribution in E3_19 and E3_5. E3_19 is depicted on the left, E3_5 is depicted on the right. (**a**) Negative charges in the β-turn regions. Negatively charged residues are depicted in stick mode in red on the backbones in ribbons. For completeness, Asp122 of E3_19 is also shown. (**b**) Positively charged residues (stick mode in blue) in the first three repeats of E3_19 and the first and fourth repeat of E3_5, respectively, are depicted on the backbones in ribbons. (**c**) Surface representations of the electrostatic potential of E3_19 and E3_5 (red = −0.5, white = 0, blue = +0.5). For orientation identically positioned ribbon representations are depicted on the left of each surface representation. This figure was prepared using MolMol.[28]

and Dr. Michael T. Stumpp for discussions. The Swiss Light Source is acknowledged for access to the PX beamline. H.K.B. was the recipient of a predoctoral fellowship of the Roche Research Foundation.

REFERENCES

1. Andrade MA, Perez-Iratxeta C, Ponting CP. Protein repeats: structures, functions, and evolution. J Struct Biol 2001;134:117–131.
2. Bork P. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? Proteins Struct Funct Genet 1993;17:363–374.
3. Groves MR, Barford D. Topological characteristics of helical repeat proteins. Curr Opin Struct Biol 1999;9:383–289.
4. Kobe B, Kajava AV. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. Trends Biochem Sci 2000;25:509–515.
5. Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D. A census of protein repeats. J Mol Biol 1999;293:151–160.
6. Sedgwick SG, Smerdon SJ. The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem Sci 1999;24:311–316.
7. D'Andrea LD, Regan L. TPR proteins: the versatile helix. Trends Biochem Sci 2003;28:655–662.
8. Kobe B, Kajava AV. The leucine-rich repeat as a protein recognition motif. Curr Opin Struct Biol 2001;11:725–732.
9. Binz HK, Amstutz P, Kohl A, Stumpp MT, Briand C, Forrer P, Grütter MG, Plückthun A. High-affinity binders selected from designed ankyrin repeat protein libraries. Nat Biotechnol 2004;22:575–582.
10. Binz HK, Stumpp MT, Forrer P, Amstutz P, Plückthun A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol 2003;332:489–503.
11. Kohl A, Binz HK, Forrer P, Stumpp MT, Plückthun A, Grütter MG. Designed to be stable: crystal structure of a consensus ankyrin repeat protein. Proc Natl Acad Sci USA 2003;100:1700–1705.
12. Main ER, Xiong Y, Cocco MJ, D'Andrea L, Regan L. Design of stable α-helical arrays from an idealized TPR motif. Structure (Camb) 2003;11:497–508.
13. Mosavi LK, Minor DL Jr, Peng Z-y. Consensus-derived structural determinants of the ankyrin repeat motif. Proc Natl Acad Sci USA 2002;99:16029–16034.
14. Mosavi LK, Peng Z-y. Structure-based substitutions for increased solubility of a designed protein. Protein Eng 2003;16:739–745.
15. Stumpp MT, Forrer P, Binz HK, Plückthun A. Designing repeat proteins: modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. J Mol Biol 2003;332:471–487.
16. Yu H, Kohl A, Binz HK, Plückthun A, Grütter MG, van Gunsteren WF. Molecular dynamics study of the stabilities of consensus designed ankyrin repeat proteins. Proteins. Forthcoming.
17. Collaborative Computational Project Number 4. Programs for protein crystallography. Acta Crystallogr D 1994;50:760–763.
18. Navaza J. AMoRe: an automated package for molecular replacement. Acta Crystallogr A 1994;50:157–163.
19. Jones TA, Zou JY, Cowan SW, Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
20. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 1998;54:905–921.
21. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ. Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr D 1999;55:247–255.
22. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
23. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52–56.
24. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
25. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 1995;8:127–134.
26. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins Struct Funct Genet 1991;11:281–296.
27. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.
28. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.
29. Batchelor AH, Piper DE, de la Brousse FC, McKnight SL, Wolberger C. The structure of GABPα/β: an ETS domain-ankyrin repeat heterodimer bound to DNA. Science 1998;279:1037–1041.