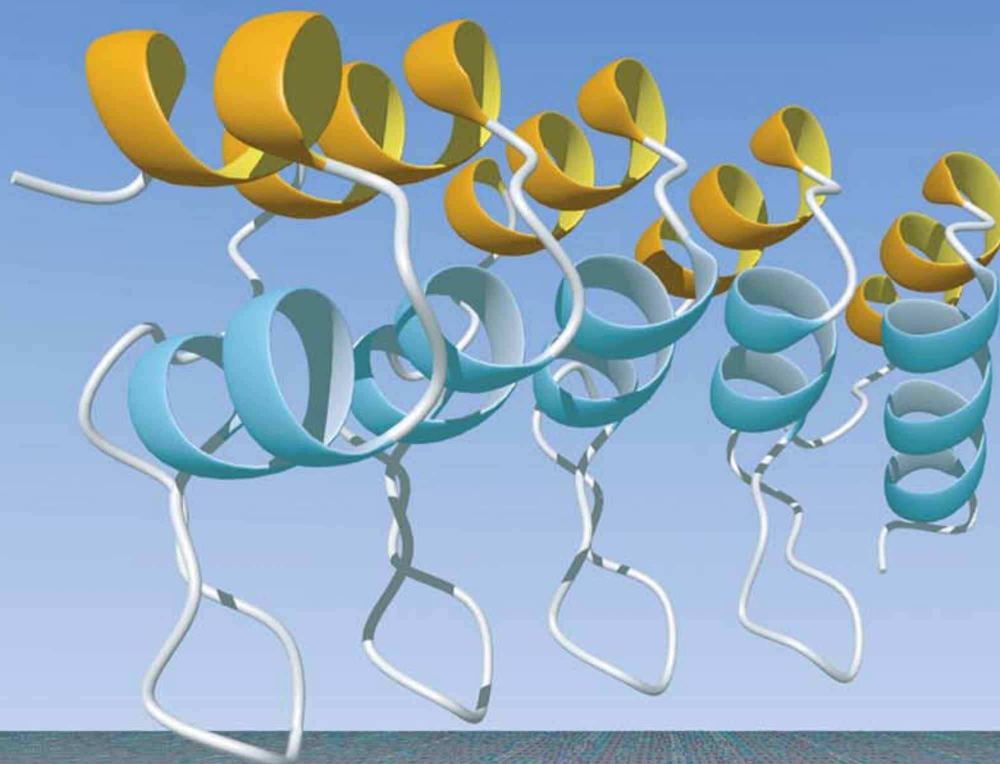
CBCHFX 5 (2) 141 – 248 (2004) · ISSN 1439-4227 · Vol. 5 · No. 2 · February 6, 2004

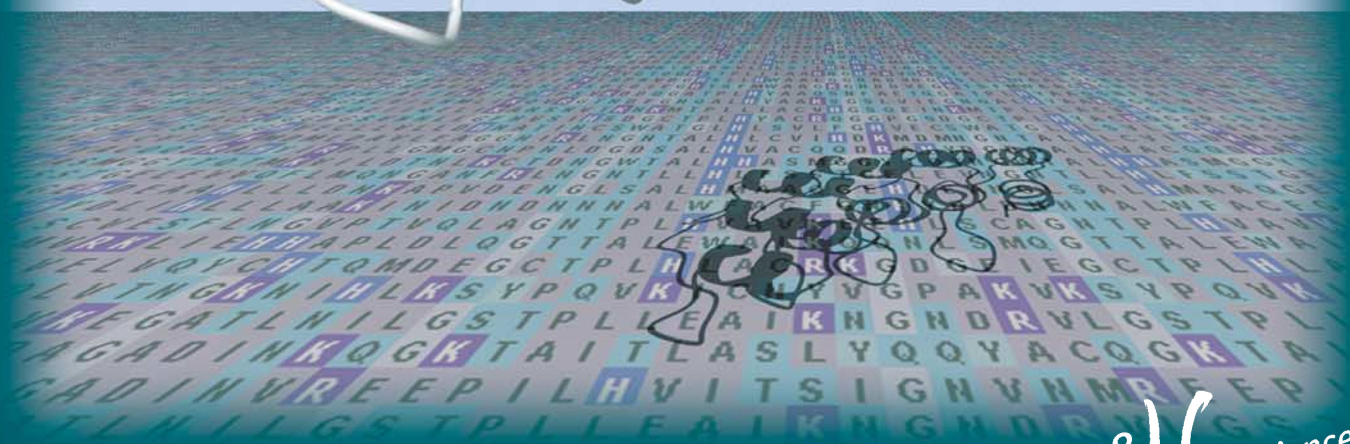
D55712

A EUROPEAN JOURNAL

CHEMBIOCHEM

OF CHEMICAL BIOLOGY





2/2004

Chemistry & Life Sciences

Review: Lipase-Specific Foldases

Concept: Consensus Design of Repeat Proteins

Minireviews: Combinatorial Approaches To Novel Proteins; Functional Changes in the Family of Type 3 Copper Proteins During Evolution; Coiled Coil Domains: Stability, Specificity, and Biological Implications

Protein Design Special Issue

Guest editors: F.-X. Schmid and R. Sterner

EUChemSoc

WILEY-VCH

Consensus Design of Repeat Proteins

Patrik Forrer, H. Kaspar Binz, Michael T. Stumpp, and Andreas Plückthun*[a]

Consensus design is a valuable protein-engineering method that is based on statistical information derived from sequence alignments of homologous proteins. Recently, consensus design was adapted to repeat proteins. We discuss the potential of this novel repeat-based approach for the design of consensus repeat proteins and repeat protein libraries and summarize recent results from such experiments.

1. Introduction

One important goal of protein engineering is to improve the biotechnological or medical efficacy of proteins by adapting their biophysical or functional properties. Besides affinity and specificity, these properties include protein expression and folding properties, as well as protein solubility and stability. However, improving these properties is not straightforward, since the underlying principles are not fully understood. Current protein engineering approaches to improve natural proteins can be grouped into three categories: i) approaches based on structure and computation, ii) those exploiting directed evolution and iii) those based on consensus design.[1–5] Here, we focus our discussion on consensus design (Table 1), bearing in mind that these approaches can all be combined.

Table 1. *Glossary*

Repeat protein	A protein containing a repeat domain.
Repeat domain	A domain composed of repeating homologous structural units (repeats), which tightly stack together forming a joined hydrophobic core. The stability of the domain is ensured by the mutual stabilization of the repeats. In most repeat domain classes, the stacking of repeats leads to an elongated architecture, but propeller-like circular architectures also exist.
Repeat	One of several repeated homologous building blocks of a repeat domain. A repeat has a well-defined topology when present in a repeat domain, but is usually unfolded on its own. Typically, a repeat of an elongated repeat domain consists of 20 to 42 amino acids and contains a characteristic amino acid sequence motif.
Capping repeat	A terminal repeat of an elongated repeat domain that shields the continuous hydrophobic core of the stacked repeats.
Consensus design	Protein design method, where a multiple alignment of a family of homologous proteins is used to calculate a consensus sequence. This consensus sequence is then compared with existing protein sequences and the differences predict which point mutations can be introduced to increase the stability of the protein. Similarly, whole proteins can be designed from a consensus sequence.
Repeat-based consensus design	Consensus design method tailored for repeat proteins. The consensus sequence of a family of homologous repeats is calculated and a self-compatible repeat is designed, which can be assembled to repeat domains.

Consensus design uses statistical analyses of sequence alignments of families of homologous proteins for protein engineering.[4, 5] The quality of such alignments depends on the number of unbiased sequences available. Thus, the reliability of consensus design strongly profits from ongoing genome projects. The idea fundamental to consensus design is that functionally important residues of proteins are more conserved than other residues; this is a direct result of diversification and selection during protein evolution. Such conserved residues include those maintaining the fold of a protein and those central to the process of protein folding and to the avoidance of aggregation. Whether residues important for biological function (e.g. binding to a particular target or catalysis) are conserved in an alignment depends on whether all of the selected family members share this function. For example, a family of enzymes will normally exhibit conserved catalytic residues, whereas the binding regions in antibodies are not conserved. Natural proteins rarely follow the consensus sequence at all structurally important positions, since natural proteins only have to be stable enough to fulfill their biological function; proteins with stabilities above a certain threshold will have no further selection advantage.[6] Thus, replacing a residue with the corresponding consensus amino acid may improve the stability or folding efficiency of a protein of interest. Furthermore, proteins designed by consensus approaches often maintain their biological function, since given residues are replaced only by amino acids that have already proven their evolutionary fitness. Consensus design therefore directly profits from millions of years of natural diversification and selection.

The efficacy of consensus design was demonstrated, for example, for the McPC603 immunoglobulin light-chain variable domain,[7] the p53 DNA binding domain,[8] the GroEL minichaperone,[9] the Abp1p SH3 domain,[10] and the tumor suppressor p16^{INK4a}.[11] In these studies 25 % to 60 % of the proteins bearing a

[a] *Dr. P. Forrer, Dipl. Biotechnol. H. K. Binz, Dipl. Natw. ETH M. T. Stumpp, Prof. Dr. A. Plückthun,*
Biochemisches Institut, Universität Zürich
Winterthurerstrasse 190, 8057 Zürich (Switzerland)
Fax: (+41) 1-635-5712
E-mail: plueckthun@bioc.unizh.ch

 DOI: 10.1002/cbic.200300762 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

single consensus designed mutation showed significantly increased thermodynamic stability, when compared to the corresponding wild-type proteins. The success rate was highest for the immunoglobulin domain (60%), where the largest set of homologous sequences was used to derive the consensus sequence. The stability gain was found to be additive for most mutations described in the various systems. For example, combining three mutations in the Abp1p SH3 domain resulted in an increase of the melting temperature (T_m) from 60°C to above 90°C and a concomitant doubling of the Gibbs free energy of unfolding (ΔG_u) to 6 kcal mol^{-1}.[10] In addition, full consensus proteins with stabilities higher than the corresponding natural proteins present in the alignment were generated.[12–15]

An important caveat is that consensus design becomes more complex when the analyzed protein family needs to be split into subfamilies with mutually exclusive residue combinations. Well studied examples are the subfamilies of antibody variable domains.[16] In such cases, consensus design has to be separately applied to each subfamily to avoid design incompatibilities. If a particular subfamily is strongly over-represented (e.g. V_H3 in the antibody V_H subfamilies), the consensus sequence of the whole family would be essentially identical to that subfamily, and all further information from other subfamilies would thus be obscured.[17]

Recently, consensus design was successfully extended to repeat proteins (Table 1), exploiting their repetitive architecture.[18–24] Here, we review this novel repeat-based consensus design (Table 1) and discuss its potential for the generation of consensus repeat proteins, the design of repeat protein libraries and recent results from such experiments.

2. The Architecture of Repeat Proteins

Repeat proteins are abundant binding molecules present in most forms of life. They are involved in innumerable biological processes, such as cell-cycle control, transcriptional regulation, innate immunity, vesicular trafficking, cell differentiation, apoptosis, cellular scaffolding, plant defense or bacterial invasion.[25–28] On the molecular level, repeat proteins mediate specific target interactions, similar to immunoglobulins. The success of repeat proteins as specific binding molecules most probably relies on their repetitive architecture, resulting in malleable and modular molecular surfaces able to mediate specific interactions (Figure 1).[26–30]

Repeat proteins contain consecutive copies of homologous structural units ("repeats", Table 1). In many repeat proteins these units stack together to form elongated, non-globular domains with a joined hydrophobic core. Elongated repeat domains are often terminated by special repeats, which we termed "capping repeats" (Table 1).[31] Such capping repeats shield the hydrophobic core of the repeat domain from the solvent. This structural arrangement within repeat domains allows, in principle, to add, exchange, or delete repeats from repeat domains without destroying their three-dimensional structure, but varying their surfaces. Individual repeats consist of framework residues important for intra- and inter-repeat

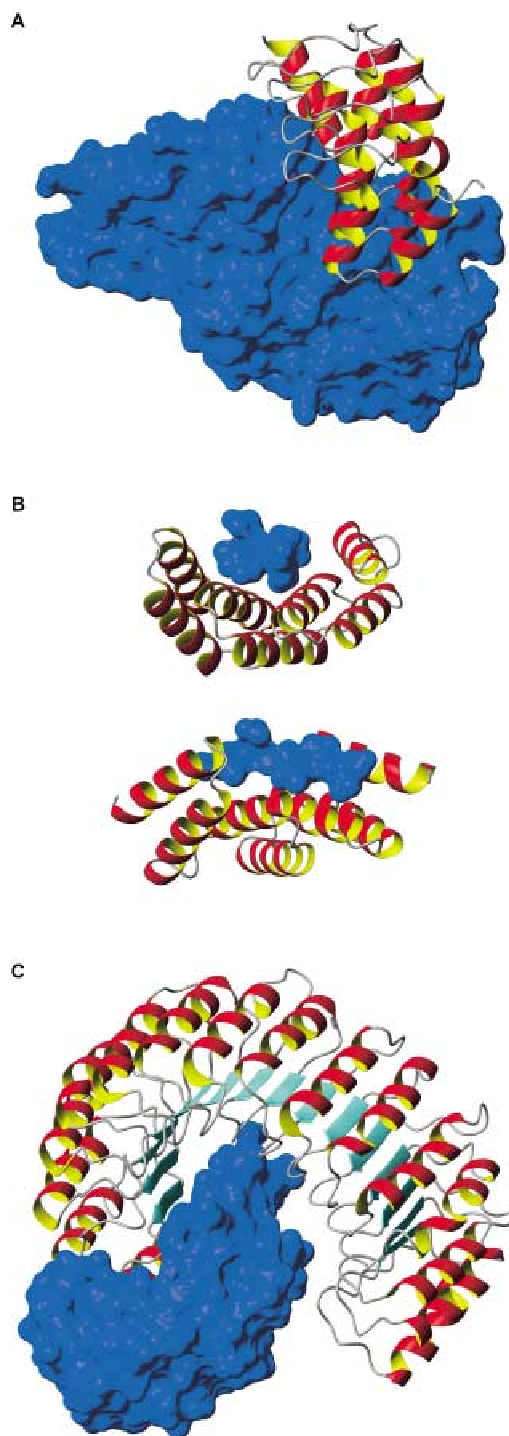


Figure 1. Target interaction of natural repeat proteins. The repeat proteins are shown as ribbon representations and the target molecules as surface representations (in blue). A) Model of the crystal structure (PDB ID: 1BLX) of the complex of the AR protein p19 bound to the cyclin-dependent kinase 6 (CDK6).[59] The five ARs of p19 together form an elongated domain that uses its concave side to bind CDK6. B) The crystal structure of the complex of the TPR1 domain of Hop bound to the C-terminal heptapeptide of the chaperone Hsp70 (PDB ID: 1ELW).[60] Two views of the complex are presented; the lower one is rotated by 45° around the vertical axis compared to the upper one. The structure shows the target peptide in an extended conformation, filling a groove in the TPR domain. C) The crystal structure of the complex of the porcine ribonuclease inhibitor (an LRR protein) with bovine ribonuclease A (PDB ID: 1DFJ).[61] The inhibitor binds the ribonuclease A with its concave surface formed by the parallel β-sheets and with the loops. A–C) The figures were prepared with MOLMOL.[62]

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

interactions and surface-exposed potential target interaction residues. The surface residues of several juxtaposed repeats together form a large potential binding surface. Typical repeats consist of 20 – 42 amino acid residues and have a well-defined topology when present in a repeat domain.[32] Each repeat consists of α-helices, β-strands, or loops in a combination typical for its class. Examples of such repeats include leucine-rich repeats (LRRs), ankyrin repeats (ARs), armadillo/HEAT repeats, and tetratricopeptide repeats (TPRs).[26, 28, 32–36] The structures of elongated repeat proteins are distinct from globular proteins: they possess an elongated shape, an intrinsic repeat-based modularity, and they lack interactions between residues very distant in sequence.

3. Repeat-Based Consensus Design

The biological importance and the particular architecture of repeat proteins make them attractive targets for protein engineering. Design and engineering of repeat proteins may help to elucidate their structural and biophysical properties, such as the dependence of stability and folding on the number of repeats, as well as the importance of key intra- and inter-repeat interactions. The vast majority of stability and folding studies so far have focused on globular proteins, leaving repeat proteins largely unexplored. On the other hand, repeat proteins are, like immunoglobulins, versatile natural scaffolds specialized for target binding. Thus, design and engineering of repeat proteins may result in novel binding molecules suitable for biotechnological or medical applications.

3.1 Concept

Recently, consensus design was adapted to the repetitive nature of repeat proteins (Table 2, Figure 2).[18–22] Instead of building repeat domains based on alignments of whole repeat domains, this novel concept involves the design of consensus repeats and their subsequent assembly into repeat domains. In this repeat-based consensus design, consensus repeats were obtained by both intra- and intermolecular sequence alignments of homologous repeats. Such intramolecular repeat alignments ensure that repeat residues present for functional reasons other than maintaining the fold, for example, specific binding functions, are averaged out. Importantly, each repeat domain may contribute several different repeat sequences to such an alignment, thereby rapidly expanding the statistical basis for the consensus design with increasing number of repeat domains available. Consensus design based on the alignment of homologous globular protein domains normally results in a consensus sequence that reflects both functionally (e.g., active site residues) and structurally important residues. In contrast, the alignment of homologous repeats, as outlined above, results in a consensus sequence that only reflects structurally important residues, that is, framework residues able to mediate conserved intra- and inter-repeat interactions. Such a conservation of inter-repeat interactions ensures the self-compatibility of consensus-designed repeats, that is, they are able to stack properly into repeat domains. Indeed, such self-compatibility of consensus-designed LRRs, ARs, and TPRs was recently demonstrated.[18–22] Interestingly, it did not seem necessary to consider the covariance of residues within the repeats separately, since certain covariations apparently dominate in nature. However, in the LRR design,[20] the consensus was derived exclusively from the subfamily of ribonuclease inhibitors to avoid any incompatibilities from mixing different LRR subfamilies. Taken together, repeat-based consensus design of repeat proteins results in designed proteins not biased by any functional constraints other than those maintaining the fold. Such repeat proteins are not only of great value for stability and folding analysis but may be an optimal

Table 2. Repeat-based consensus design studies

Reference	Names of designed proteins (PDB ID)	Fold (number of repeats)	Number of repeat sequences considered	Design characteristics	Key results
Mosavi et al.[18]	1ANK 2ANK 3ANK (1NOQ) 4ANK (1NOR)	ankyrin repeat (1-4)	4400	sequence-based, no capping	stable regular AR domains
Binz et al.[21] and Kohl et al.[19]	N2C N3C (1MJO) N4C	ankyrin repeat (4-6)	2220	sequence-based, structural refinement, capping repeats,[a] library	stable regular AR domains, soluble expression, domains with diversified surfaces
Main et al.[22]	CTPR1 CTPR2 (1NA3) CTPR3 (1NA0)	tetratricopeptide repeat (1.5-3.5)	1837	sequence based, capping features[b]	stable regular TPR domains
Stumpp et al.[20]	N3C N4C N5C N6C	leucine-rich repeat (8-14)	28[c]	sequence-based, structural refinement, capping repeats,[a] library	soluble expression, domains with diversified surfaces

[a] Designed capping repeats terminate the repeat domain. [b] A helix nucleation sequence was added to the first helix at the N-terminus and a solvating helix to the C-terminus. [c] 28 "double repeats", each consisting of an A- and B-type LRR, typical for mammalian ribonuclease inhibitors, were considered.

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

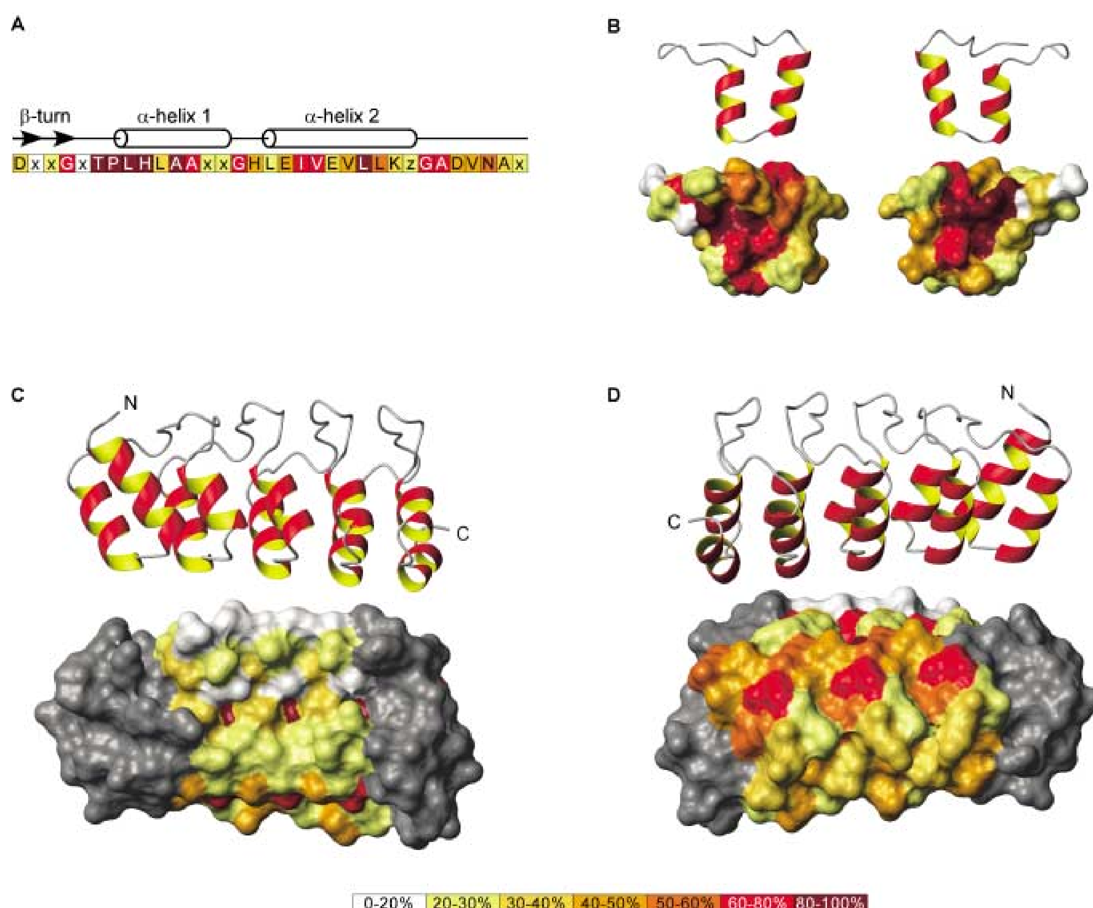


Figure 2. Consensus analysis of AR proteins. The residue conservation of AR proteins, as determined by Binz et al.,[21] is illustrated by using the X-ray structure of the designed AR protein E3_5 (PDB ID: 1MJ0).[19] E3_5 consists of three consensus-designed repeats flanked by capping repeats. The consensus repeats are based on the repeat sequence motif of (A). The percentage of conservation of each amino acid position is color coded as indicated at the bottom of the figure. A) The AR consensus sequence motif as designed by Binz et al.[21] x: Any amino acid but not G, C, or P. z: Any of the amino acids H, N, Y. The motif is also colored according to the level of amino acid conservation. B) Illustration of the residue conservation in a surface representation of an AR. The middle repeat of E3_5 is presented in a lateral view from the C terminus (left) and from the N terminus (right) showing the conserved hydrophobic core and less conserved surface exposed residues. C) and D) Surface representations of E3_5. The concave surface of the AR domain (C) is less conserved than the opposite side (D) and was used for randomization in the AR protein library design.[21] The capping repeats (shown in gray) differ from central repeats and were thus excluded from the consensus analysis. B)–D) For orientation, corresponding ribbon representations are shown on top of each surface representation. The figures were prepared with MOLMOL.[62]

starting point for the design of novel binding molecules (see below).

3.2. Consensus repeat proteins

Whereas small globular proteins have been studied extensively in terms of folding and stability, only a few natural repeat proteins have been investigated.[37–44] Nevertheless, the structural peculiarities of repeat proteins ask for a more fundamental analysis. Especially, the influence of the number of repeats on the folding and stability of repeat domains may give deeper insights into the cooperativity of the inter-repeat interactions. Although repeats of natural repeat proteins are structurally homologous, evolutionary drift[45] is thought to have produced substantial sequence variations. The resulting sequence heterogeneity complicates the interpretation of the fundamental biophysical properties of repeat protein architectures.[24, 39] Therefore, repeat proteins obtained by repeat-based consensus design are more suitable for the analysis of these fundamental properties. As outlined above, such designed repeat proteins are free of any constraints other than scaffold functions.

Mosavi et al. (Table 2) designed AR polypeptides containing one to four identical repeats.[18] Whereas the obtained one- and two-repeat proteins were mainly unfolded, the obtained three- and four-repeat proteins were well-folded monomers upon refolding as indicated by gel filtration, CD, and NMR measurements. In addition, thermal denaturation of the larger proteins was monitored by CD measurements at 222 nm, which showed reversible, cooperative thermal transitions between pH 4 and 5. The determined midpoints of transitions were 69.4 °C and 81.3 °C for the three- and four-repeat protein, respectively. Thus, their results show a clear increase in stability as the repeat number increases. X-ray crystallographic analysis of these proteins demonstrated that they adopt a very regular, tightly packed AR fold, proving the success of repeat-based consensus design.

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Binz et al. (Table 2, Figure 2) designed consensus AR proteins with fixed framework and randomized interaction residues to generate a regular scaffold displaying variable molecular surfaces.[19, 21] Thereby, designed AR proteins were obtained that could be expressed at levels of 200 mg L^{-1} soluble protein in *Escherichia coli* using normal shake flasks. Melting temperatures from 66 °C to above 85 °C (CD measurements) and thermodynamic stabilities between 9 and 21 kcal mol^{-1} (denaturant-induced equilibrium unfolding) were determined. Thus, these proteins have considerably improved stabilities and expression yields when compared to natural AR proteins of the same length. In addition, the crystal structure of one of these designed AR proteins provided insight into a very regular and tightly packed AR fold. Again, designed AR proteins with more repeats exhibited higher stabilities.

Main et al. (Table 2) designed consensus TRP proteins containing one to three identical repeats.[22] These constructs also contained three additional N-terminal residues to cap the N-terminal α-helix and a polar C-terminal helix to cap the C-terminal repeat. Analytical ultracentrifugation and analytical gel filtration showed that all three proteins were monomeric in solution over a wide concentration range (from 10 μM to 2 mM). Far-UV CD spectra indicated high α-helical content. Furthermore, the proteins exhibited cooperative thermal unfolding transitions when followed by CD, with midpoints of thermal denaturations of 49 °C, 74 °C, and 83 °C for their one-, two- and three-repeat proteins, respectively. Thus, increasing the number of repeats resulted in a concomitant increase in the melting temperature. The X-ray structures of the two- and three-repeat constructs closely match that of natural TPR proteins; this confirms the success of the design strategy.

Stumpp et al. (Table 2) designed consensus LRR proteins with randomized surfaces. These proteins were well expressed, monomeric, and showed α-helical CD spectra.[20] Interestingly, the equilibrium folding behavior of these designed repeat proteins could not be described by a two-state approximation and appears to follow a more complicated model. In this study, the consensus analysis was undertaken for the mammalian ribonuclease inhibitor family, which contains some of the tightest binding molecules known.[46] At the time of this consensus analysis, the sequences of only 28 homologous repeats were known; this led to the smallest dataset used for repeat-based consensus design of the examples mentioned. In contrast to the other repeat protein families, the LRR protein family can be divided into at least six LRR subfamilies,[47] and consensus design preferably relies only on proteins from one subfamily. Earlier reports described LRR consensus sequences averaging over all known LRR proteins,[34] which would obscure the sequence pattern and consequently would not be useful for consensus design approaches.

An interesting design aspect of the consensus-designed AR and LRR protein libraries[19–21] is that they are terminated by capping repeats, which may be beneficial for domain folding and stability through shielding of the hydrophobic core of these domains from the solvent. AR and LRR proteins with capping repeats and identical consensus repeats have also been generated to study their biophysical properties as a function of the repeat number.[48]

All in all, repeat-based consensus design of repeat proteins has proven to be a powerful tool for obtaining consensus repeat proteins of the AR, TPR, and LRR folds; this indicates that it is most probably also applicable to other repeat proteins. The future analysis of such designed repeat proteins will certainly provide new insights into their stability and folding. From the available data, it is already clear that repeat proteins need a minimum number of repeats to be stable. Moreover, it is likely that the stability of all repeat proteins increases with increasing repeat numbers. In addition, consensus-designed repeat proteins seem to be more stable than their natural counterparts; this underscores the general benefit of consensus design to improve the biophysical properties of proteins.

3.3. Repeat protein libraries

Man-made polypeptide libraries have become indispensable sources for specific binding molecules in research and biomedical applications. Antibody fragments[49, 50] are currently by far the most widely used binding molecules. Nevertheless, antibody fragments can have limitations in expression yield and stability, depending on the antibody sequence. One particularly interesting application is to express antibodies intracellularly to inhibit cellular functions. Yet, under the reducing conditions of the cytoplasm, the disulfide bonds of the antibody cannot form, thus destabilizing the antibody.[51]

One way to deal with this limitation is to engineer antibodies for higher stability by using consensus-based, structure-based, or evolutionary approaches.[52–54] Another way to circumvent the limitations of antibodies is to develop novel binding molecules based on other protein architectures.[55–57] Such molecules should combine the high affinity and specificity of antibody fragments with efficient folding and expression properties as well as with high thermodynamic stability under both oxidizing and reducing conditions. Forrer et al. envisioned that repeat proteins may be an attractive alternative to antibodies because, next to antibodies, repeat proteins constitute the most abundant natural protein classes specialized for binding.[31] Unlike antibodies, repeat proteins occur intra- and extracellularly, and their success as binding molecules most likely relies on their modular architecture. Repeat proteins differ from almost all other binding molecules by having a modular interaction surface whose size can be simply adapted by varying the number of repeats. Based on this concept of modularity, a novel strategy to generate combinatorial libraries of repeat proteins of variable length with randomized target interaction residues was developed that led to highly diversified molecular surfaces.[31] The key steps of this strategy are the use of repeat-based consensus design to generate a diverse set of compatible repeat modules and the subsequent assembly of these modules into repeat protein libraries between two capping repeats. Randomly chosen members from such libraries indeed showed good expression, folding, and stability properties, exceeding those of corresponding natural repeat proteins, while displaying variable surface residues (see above and Table 2).[19–21] From the AR

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

protein libraries,[21] we isolated specific binding molecules against several globular proteins with affinities in the low nanomolar range by using ribosome display.[58] Their very high expression yield and high stabilities make them attractive binding molecules for various biotechnological applications.

The use of repeat-based consensus design to build repeat protein libraries has several advantages over other engineering approaches: i) In principle, only primary structural data of homologous repeats are needed to design consensus repeats with randomized positions. Nevertheless, tertiary structure information may be very helpful for such a design approach, particularly when deciding on less-conserved positions.[19–21] ii) Consensus design ensures that all natural binding functions of the input repeat proteins are averaged out, leaving an "inert" scaffold suitable to display novel binding specificities. iii) This approach directs the focus on positions that can be randomized due to their low conservation, even though, again, structural information from natural proteins may be very helpful.[19–21] iv) Consensus design ensures the self-compatibility of repeats designed in this manner, which is the key for their proper assembly in repeat domains. (v) Finally, repeat-based consensus design seems to lead to libraries in which the vast majority of members possess high thermodynamic stabilities, even though around 20% of the residues are randomized.[19–21] The high stability is a very desirable property as it makes the library members more tolerant to destabilizing, but functionally beneficial mutations, which may occur during their in vitro evolution.

Conclusions and Perspectives

Recently, repeat proteins of the AR, LRR, and TPR families have been engineered by a novel repeat-based consensus design approach allowing the generation of consensus repeat proteins or combinatorial repeat protein libraries.[18–22] This approach strongly profited from the large sequence datasets available for repeat proteins. We think that this approach has three major implications for protein engineering. First, consensus repeat proteins have given the first insights into fundamental properties related to their repetitive architecture. The different architecture of repeat proteins compared with globular proteins makes them a novel and important focus for protein folding and stability studies. In this regard, the dependence of their biophysical properties on the number of repeats is of special interest. Second, consensus repeat proteins may constitute attractive model systems to study sequence–structure relationships, as their primary and tertiary structures are much simpler than those of globular proteins. Only a small polypeptide stretch, that is, a repeat, has to be designed and all particular design decisions, whether they have positive or negative effects, will potentiate through the assembly of the repeats into domains. Finally, repeat-based consensus design is an attractive tool for the generation of novel binding molecules.[19–21, 31] There, consensus design not only ensures high stability, but also helps to introduce the needed diversity.

Acknowledgements

We thank Patrick Amstutz, Armela Hatkic, Svava Wetzel, and Christian Zahnd for valuable discussions. H.K.B. was the recipient of a predoctoral fellowship of the Roche Research Foundation. M.T.S. was the recipient of a FCI and BMBF predoctoral fellowship. This project was supported by the NCCR Structural Biology and the Swiss Cancer Research grant KFS 1055-09-2000.

Keywords: combinatorial library · consensus design · consensus sequence · protein design · protein engineering

[1] M. B. Tobin, C. Gustafsson, G. W. Huisman, *Curr. Opin. Struct. Biol.* **2000**, *10*, 421.
[2] F. H. Arnold, P. L. Wintrode, K. Miyazaki, A. Gershenson, *Trends Biochem. Sci.* **2001**, *26*, 100.
[3] A. Plückthun, C. Schaffitzel, J. Hanes, L. Jermutus, *Adv. Protein Chem.* **2000**, *55*, 367.
[4] M. Lehmann, M. Wyss, *Curr. Opin. Biotechnol.* **2001**, *12*, 371.
[5] B. van den Burg, V. G. H. Eijsink, *Curr. Opin. Biotechnol.* **2002**, *13*, 333.
[6] D. M. Taverna, R. A. Goldstein, *Proteins* **2002**, *46*, 105.
[7] B. Steipe, B. Schiller, A. Plückthun, S. Steinbacher, *J. Mol. Biol.* **1994**, *240*, 188.
[8] P. V. Nikolova, J. Henckel, D. P. Lane, A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14675.
[9] Q. Wang, A. M. Buckle, N. W. Foster, C. M. Johnson, A. R. Fersht, *Protein Sci.* **1999**, *8*, 2186.
[10] A. Rath, A. R. Davidson, *Protein Sci.* **2000**, *9*, 2457.
[11] T. J. Cammett, L. Luo, Z. Peng, *J. Mol. Biol.* **2003**, *327*, 285.
[12] A. Knappik, L. Ge, A. Honegger, P. Pack, M. Fischer, G. Wellnhofer, A. Hoess, J. Wölle, A. Plückthun, B. Virnekäs, *J. Mol. Biol.* **2000**, *296*, 57.
[13] M. Lehmann, D. Kostrewa, M. Wyss, R. Brugger, A. D'Arcy, L. Pasamontes, A. P. G. M. van Loon, *Protein Eng.* **2000**, *13*, 49.
[14] M. Lehmann, C. Loch, A. Middendorf, D. Studer, S. F. Lassen, L. Pasamontes, A. P. G. M. van Loon, M. Wyss, *Protein Eng.* **2002**, *15*, 403.
[15] M. J. Macias, V. Gervais, C. Civera, H. Oschkinat, *Nat. Struct. Biol.* **2000**, *7*, 375.
[16] S. Ewert, T. Huber, A. Honegger, A. Plückthun, *J. Mol. Biol.* **2003**, *325*, 531.
[17] M. Visintin, G. Settanni, A. Maritan, S. Graziosi, J. D. Marks, A. Cattaneo, *J. Mol. Biol.* **2002**, *317*, 73.
[18] L. K. Mosavi, D. L. Minor, Jr., Z. Peng, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16029.
[19] A. Kohl, H. K. Binz, P. Forrer, M. T. Stumpp, A. Plückthun, M. G. Grütter, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1700.
[20] M. T. Stumpp, P. Forrer, H. K. Binz, A. Plückthun, *J. Mol. Biol.* **2003**, *332*, 471.
[21] H. K. Binz, M. T. Stumpp, P. Forrer, P. Amstutz, A. Plückthun, *J. Mol. Biol.* **2003**, *332*, 489.
[22] E. R. G. Main, Y. Xiong, M. J. Cocco, L. D'Andrea, L. Regan, *Structure* **2003**, *11*, 497.
[23] E. R. G. Main, S. E. Jackson, L. Regan, *Curr. Opin. Struct. Biol.* **2003**, *13*, 1.
[24] K. W. Tripp, D. Barrick, *Structure* **2003**, *11*, 486.
[25] D. A. Jones, J. D. G. Jones, *Adv. Bot. Res.* **1997**, *24*, 89.
[26] M. A. Andrade, C. Perez-Iratxeta, C. P. Ponting, *J. Struct. Biol.* **2001**, *134*, 117.
[27] B. Kobe, A. V. Kajava, *Curr. Opin. Struct. Biol.* **2001**, *11*, 725.
[28] S. G. Sedgwick, S. J. Smerdon, *Trends Biochem. Sci.* **1999**, *24*, 311.
[29] G. L. Blatch, M. Lässle, *BioEssays* **1999**, *21*, 932.
[30] X. Wang, J. McLachlan, P. D. Zamore, T. M. Tanaka Hall, *Cell* **2002**, *110*, 501.
[31] P. Forrer, M. T. Stumpp, H. K. Binz, A. Plückthun, *FEBS Lett.* **2003**, *539*, 2.
[32] B. Kobe, A. V. Kajava, *Trends Biochem. Sci.* **2000**, *25*, 509.
[33] M. R. Groves, D. Barford, *Curr. Opin. Struct. Biol.* **1999**, *9*, 383.
[34] B. Kobe, J. Deisenhofer, *Trends Biochem. Sci.* **1994**, *19*, 415.
[35] B. Kobe, *Nat. Struct. Biol.* **1996**, *3*, 977.
[36] M. A. Andrade, C. Petosa, S. I. O'Donoghue, C. W. Müller, P. Bork, *J. Mol. Biol.* **2001**, *309*, 1.
[37] C. Marchetti Bradley, D. Barrick, *J. Mol. Biol.* **2002**, *324*, 373.
[38] L. K. Mosavi, S. Williams, Z. Peng, *J. Mol. Biol.* **2002**, *320*, 165.

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[39] K. S. Tang, A. R. Fersht, L. S. Itzhaki, *Structure* **2003**, *11*, 67.
[40] M. Zeeb, H. Rösner, W. Zeslawski, D. Canet, T. A. Holak, J. Balbach, *J. Mol. Biol.* **2002**, *315*, 447.
[41] B. Zhang, Z. Peng, *J. Mol. Biol.* **2000**, *299*, 1121.
[42] K. S. Tang, B. J. Guralnick, W. K. Wang, A. R. Fersht, L. S. Itzhaki, *J. Mol. Biol.* **1999**, *285*, 1869.
[43] M. E. Zweifel, D. Barrick, *Biochemistry* **2001**, *40*, 14357.
[44] M. E. Zweifel, D. Barrick, *Biochemistry* **2001**, *40*, 14344.
[45] M. Kimura, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5969.
[46] F. S. Lee, B. L. Vallee, *Biochem. Biophys. Res. Commun.* **1989**, *160*, 115.
[47] A. V. Kajava, *J. Mol. Biol.* **1998**, *277*, 519.
[48] M. T. Stumpp, A. Hatkic, S. Wetzel, H. K. Binz, A. Plückthun, unpublished results.
[49] A. Plückthun, A. Krebber, C. Krebber, U. Horn, U. Knüpfer, R. Wenderoth, L. Nieba, K. Proba, D. Riesenberg in *Antibody Engineering* (Eds: J. McCafferty, H. R. Hoogenboom, D. J. Chiswell), Oxford University Press, New York, **1996**, pp. 203.
[50] G. Winter, *FEBS Lett.* **1998**, *430*, 92.
[51] A. Cattaneo, S. Biocca, *Trends Biotechnol.* **1999**, *17*, 115.
[52] A. Wörn, A. Plückthun, *J. Mol. Biol.* **2001**, *305*, 989.
[53] S. Ewert, A. Honegger, A. Plückthun, *Biochemistry* **2003**, *42*, 1517.
[54] L. Jermutus, A. Honegger, F. Schwesinger, J. Hanes, A. Plückthun, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 75.
[55] A. Skerra, *J. Mol. Recognit.* **2000**, *13*, 167.
[56] R. C. Ladner, A. C. Ley, *Curr. Opin. Biotechnol.* **2001**, *12*, 406.
[57] P.-Å. Nygren, M. Uhlén, *Curr. Opin. Struct. Biol.* **1997**, *7*, 463.
[58] H. K. Binz, P. Amstutz, C. Zahnd, E. Prenosil, M. T. Stumpp, P. Forrer, A. Plückthun, unpublished results.
[59] D. H. Brotherton, V. Dhanaraj, S. Wick, L. Brizuela, P. J. Domaille, E. Volyanik, X. Xu, E. Parisini, B. O. Smith, S. J. Archer, M. Serrano, S. L. Brenner, T. L. Blundell, E. D. Laue, *Nature* **1998**, *395*, 244.
[60] C. Scheufler, A. Brinker, G. Bourenkov, S. Pegoraro, L. Moroder, H. Bartunik, F. U. Hartl, I. Moarefi, *Cell* **2000**, *101*, 199.
[61] B. Kobe, J. Deisenhofer, *J. Mol. Biol.* **1996**, *264*, 1028.
[62] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51.

Received: September 4, 2003 [C762]

 © 2004 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim