PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

The NMR solution structure of the artificial protein M7 matches the computationally designed model

Claudius Stordeur,[1] Roman Dallüge,[1] Olaf Birkenmeier,[1] Hans Wienk,[2] Rainer Rudolph,[1] Christian Lange,[1]* and Christian Lücke[3]*

[1] Institute of Biochemistry and Biotechnology, Martin-Luther-University Halle-Wittenberg, D-06099 Halle/Saale, Germany

[2] Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, 3584 CH Utrecht, The Netherlands

[3] Max Planck Research Unit for Enzymology of Protein Folding, D-06120 Halle/Saale, Germany

***Key words:** tetrapeptide fragment-based protein design algorithm; artificial fold; de novo protein.*

INTRODUCTION

Rapid progress has been made in the field of protein design over the past couple of years. Computational design algorithms have been used to (i) stabilize naturally occurring proteins,[1,2] (ii) increase the solubility of membrane proteins,[3] and (iii) create de novo proteins that feature previously unknown structural folds.[4] Hence, computational protein design opens a multitude of possibilities to study the complex processes of protein folding and to possibly create new structures with novel functional properties. The first completely de novo designed protein TOP7 (PDB 1QYS), for example, provided the opportunity to explore the energy landscape of a protein structure that was never exposed to natural selection pressure.[5] The long-term goal of this research field is the ability to design proteins that perform the desired functions and exhibit high stability for possible application in industry and therapy.

A major objective in the development of protein design is to create algorithms that are as simple as possible and need little processor time, yet are able to generate proteins that match a target structure with high accuracy. The recently published tetrapeptide fragment-based design algorithm[6] is such a simple method, using statistical data of backbone conformations that are derived from tetrapeptide segments of full-length proteins. The algorithm was tested with the novel fold of the de novo protein TOP7[4] as target structure. Several nonhomologous amino acid sequences created with this new algorithm led to protein variants that exhibited cooperative folding and high thermodynamic stability.[6] The most stable of these variants, named M7 [Fig. 1(A)], featured cooperative folding/unfolding transitions typical of naturally occurring proteins and circular dichroism spectra that indicated a well-ordered tertiary structure.

It remained unclear, however, to what degree M7 matches the designed model, which is highly homologous to the TOP7 target structure (backbone RMSD = 0.13 Å). In the present study, the structure of M7 was solved with multidimensional NMR, revealing a well-ordered artificial protein that closely resembles the predicted fold (1.24 ± 0.10 Å average backbone RMSD). This result demonstrates the high precision of the structure predic-

Abbreviations: HSQC, heteronuclear single quantum correlation; HTQC, heteronuclear triple quantum correlation; NOE, nuclear Overhauser effect; NOESY, nuclear overhauser enhancement spectroscopy; RMSD, root-mean-square deviation; TOCSY, total correlation spectroscopy.

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: EU-NMR; Grant number: RII3-026145.

*Correspondence to: C. Lange, Institut für Biochemie und Biotechnologie, Martin-Luther-Universität Halle-Wittenberg, Kurt-Mothes-Str. 3, D-06099 Halle/Saale, Germany. E-mail: christian.lange@biochemtech.uni-halle.de; or C. Lücke, Max-Planck-Forschungsstelle für Enzymologie der Proteinfaltung, Weinbergweg 22, D-06120 Halle/Saale, Germany. E-mail: luecke@enzyme-halle.mpg.de

Received 11 January 2008; Revised 17 March 2008; Accepted 25 March 2008
Published online 22 May 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22107

© 2008 WILEY-LISS, INC.

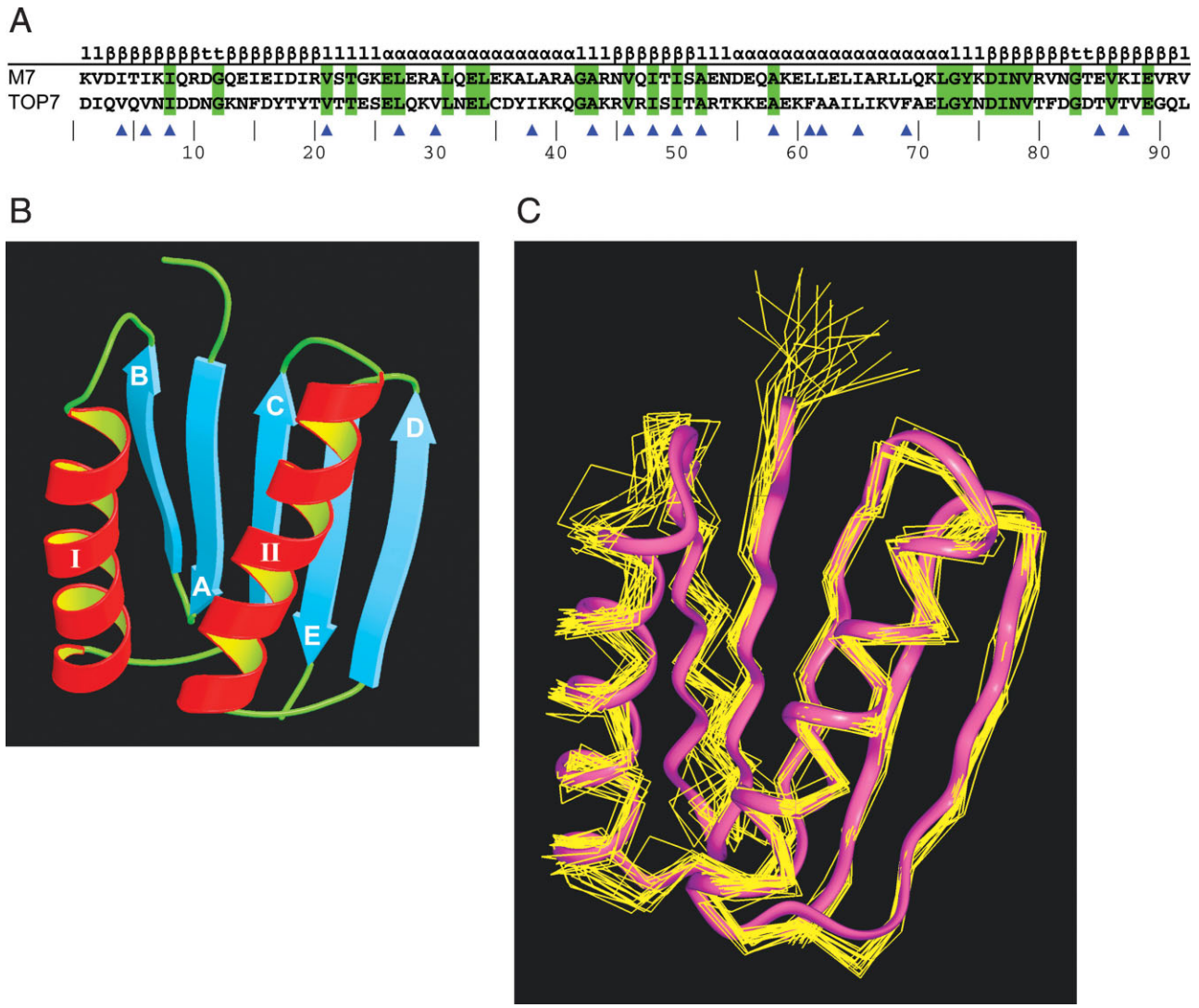

Figure 1

*(**A**) Alignment of the protein sequences of M7 and TOP7. Residues identical to the sequence of TOP7 are shaded in green. Secondary structure elements of M7 are indicated above the line as α-helix (α), β-strand (β), regular β-turn (t) or loop (l). Blue triangles mark the position of residues with greater than 40% occluded surface in TOP7.[7] (**B**) Ribbon diagram highlighting the secondary structure elements of the M7 solution structure. The five antiparallel β-strands (A–E) are indicated as cyan arrows, while the two α-helices (I and II) are shown in red/yellow. (**C**) Backbone trace superposition of the final 20 energy-minimized M7 conformers (yellow lines) with the designed structural model (magenta ribbon).*

tion by the tetrapeptide fragment-based design algorithm.

MATERIALS AND METHODS

NMR spectroscopy

The M7 protein was produced as previously described,[6] except that ^{15}N-enriched M7 was obtained by expression in M9 minimal medium. The NMR samples contained either nonlabeled or ^{15}N-enriched M7 (1.0–1.3 m*M*) in 25 m*M* sodium phosphate buffer (pH 6.5, 0.05% sodium azide, 5% D_2O). For hydrogen/deuterium exchange measurements, the samples were dissolved in perdeuterated buffer. NMR data were collected at 25°C with a DRX500 spectrometer (Bruker, Rheinstetten, Germany), operating at 500.13 MHz proton resonance frequency and equipped with a 5-mm inverse triple-resonance (^{1}H/^{13}C/^{15}N) probe that had XYZ-gradient capability. Homonuclear two-dimensional (2D) spectra (TOCSY and NOESY) as well as ^{15}N-edited multidimensional spectra (2D HSQC, 2D HTQC, 3D TOCSY-HSQC, and 3D NOESY-HSQC) were acquired and processed as previously described.[8] ^{15}N relaxation measurements were performed at 500.13 MHz as described elsewhere.[9] R1 rates resulted from fitting of signal intensities from a series of

10 experiments with relaxation delays ranging from 100 to 1200 ms. R2 rates resulted from 11 experiments recorded with delays ranging from 8 to 192 ms. $^{15}N\{^{1}H\}$-heteronuclear NOE values were calculated as the ratio of signal intensities from experiments with and without proton saturation for 3 s. Model-free analysis[10] was performed assuming isotropic rotational diffusion.

Structure determination

The sequence-specific ^{1}H and ^{15}N resonance assignments were derived via NOE connectivities according to the classical assignment strategy.[11] NOE-derived distance restraints were determined from 2D and 3D NOESY spectra employing an internal distance calibration. Automated assignments of the NOEs, based on chemical shifts only, were obtained with the program *nmr2st*.[12]

The structures were computed, as described elsewhere,[8] on a Silicon Graphics O_2 workstation with the program DYANA 1.5[13] using 1031 meaningful distance restraints that were obtained out of a total of 1477 NOE assignments. Subsequent restrained energy minimization was performed with the DISCOVER module of the INSIGHT 2000 software package (Accelrys, San Diego, CA). The 20 final selected M7 conformers were analyzed with PROCHECK-NMR.[14]

RESULTS AND DISCUSSION

The resonance assignment of M7 was completed, except for residues Gly-3 and Ser-2 that remained after cleavage of the His_6-tag. The list of chemical shift values has been deposited at the BioMagResBank database under accession number BMRB-15478. Subsequent structure calculations were performed with the assigned residues His-1 to Val92.

The final ensemble of 20 energy-minimized M7 conformers has been deposited at the RCSB database under PDB ID code 2JVF. It features an average backbone RMSD for the nonterminal residues Asp3-Arg91 of 1.09 ± 0.14 Å and only few violations of the experimental constraints (Table I). The overall structure of M7 [Fig. 1(B)] consists of a five-stranded antiparallel β-sheet, arranged in a +1, −2, −2, +1 topology, and two nearly parallel α-helices. The five β-strands comprise residues Asp3-Arg10 (A), Gln13-Arg20 (B), Asn45-Ser51 (C), Asp76-Asn82 (D), and Glu85-Arg91 (E), whereas the two α-helices encompass the segments Glu26-Ala41 (I) and Asp55-Leu72 (II). Excluding the more flexible turn regions, the average backbone RMSD of the secondary structure elements drops down to 0.91 ± 0.15 Å. Analysis of the backbone dihedral angles ϕ and ψ furthermore showed that less than 1% of all non-glycine/non-proline residues in M7 fall within the disallowed regions of conformational space (Table I).

Interestingly, the N-terminal half of the M7 structure (i.e., residues Asp3-Gly42 comprising β-strands A and B as well as α-helix I) displays a significantly lower structural precision (average backbone RMSD = 1.26 ± 0.21 Å), compared to the C-terminal half (i.e., residues Arg44-Arg91 comprising β-strands C, D, and E as well as α-helix II) with an average backbone RMSD of 0.70 ± 0.12 Å [Fig. 1(C)]. Moreover, H/D exchange data also demonstrated that the majority of M7 residues showing slowed amide proton exchange resides in the C-terminal half of the protein (Ile48, Thr49, Ile50, Asp76, and Val90 in the β-sheet; Leu61, Leu62, Leu64, Ile65, Ala66, Leu68, Leu69, and Leu72 in α-helix II; Ile77 in the turn αII-βD, possibly H-bonded to the backbone oxygen of Lys75), whereas in the N-terminal half no slowed exchange was observed at all in α-helix I and only for two residues in the β-sheet (Ile4 and Ile19). Hence it appears that the C-terminal half of M7 is more rigid, as evidenced by a stronger hydrogen bond network. A similar correlation was also reported for TOP7.[4,15]

The increased conformational divergence observed in the N-terminal half of M7 [Fig. 1(C)] is most pronounced in the loop preceding α-helix I, where it could be a result of the limited number of structurally relevant NOEs found for residues Ser22, Thr23, and Gly24. On the other hand, this lack of NOE connectivities is in part due to considerable line-broadening observed for the backbone amide signals of Ser22 and especially Gly24. Such line-broadening effects, which are indicative of differences in backbone dynamics, were furthermore found for residues Asp11 and Gly12 (β-turn between β-strands A and B) as well as Ala41 and Gly42 (end of α-helix I), that is, residues that all reside in the N-terminal half of the protein structure. Preliminary analysis of ^{15}N relaxation data [see supplementary material] indicates that these particular loops/turns all feature reduced generalized order parameters S^2, in addition to the N-terminus (Met0 and Lys1) and the turn region βD-βE (Thr84 and Val86). The origin of this increased backbone mobility mostly within the N-terminal

Table I
Structural Statistics of the Final 20 Energy-Minimized M7 Conformers

NOE-derived distance restraints	1031
Intraresidual ($i = j$)	223
Sequential ($\|i - j\| = 1$)	310
Medium-range ($1 < \|i - j\| \leq 4$)	280
Long-range ($\|i - j\| > 4$)	218
Total number of restraint violations >0.3 Å	0
Total number of restraint violations >0.2 Å	5
Maximal restraint violation (Å)	0.3
Ramachandran plot (%)	
Most favored regions	81.6
Additionally allowed regions	16.6
Generously allowed regions	1.1
Disallowed regions	0.7
Structural precision (Å)	
Backbone atom[a] RMSD (residues Asp3-Arg91)	1.09 ± 0.14
Heavy atom RMSD (residues Asp3-Arg91)	1.92 ± 0.12
Backbone atom[a] RMSD (secondary structure elements)	0.91 ± 0.15
Heavy atom RMSD (secondary structure elements)	1.84 ± 0.14

[a]N, C^{α}, C′, and O.

half of the M7-fold still needs to be determined. Possibly, the lack of stabilizing residues at the N-terminal cap[16] of α-helix I could play a role.

The comparison of the NMR ensemble with the previously designed structural model of M7[6] revealed a very good agreement in the overall fold [Fig. 1(C)]. The average backbone RMSD between the model and the 20 conformers amounts to 1.24 ± 0.10 Å for the segment Asp3-Arg91, which is almost within the structural precision of the NMR ensemble itself. Minor differences in the secondary structure elements were observed only in β-strand A, β-strand D, and α-helix I (all one residue longer at the N-terminal end in the model) as well as β-strand B (one residue longer at the C-terminal end in the model). Moreover, a stabilizing salt-bridge, which had been predicted by the model between Glu35 (α-helix I) and Lys71 (α-helix II), was found in all but one of the 20 NMR conformers.

In summary, it can be stated that the amino acid sequence, which was created by the tetrapeptide fragment-based design algorithm, produced an overall fold that matches the target structure very closely. Future studies with alternative structural targets will have to show (i) how well this tetrapeptide fragment-based design approach performs in other cases, and (ii) whether functionally active proteins can be created by this method. The present study, however, clearly demonstrates that the structural information encoded in the tetrapeptide fragments used in this relatively simple approach is sufficient to yield a high precision in structure design.

ACKNOWLEDGMENTS

The authors thank Monika Seidel (Max Planck Research Unit for Enzymology of Protein Folding) for technical assistance.

REFERENCES

1. Korkegian A, Black ME, Baker D, Stoddard BL. Computational thermostabilization of an enzyme. Science 2005;308:857–860.
2. Dantas G, Corrent C, Reichow SL, Havranek JJ, Eletr ZM, Isern NG, Kuhlman B, Varani G, Merritt EA, Baker D. High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. J Mol Biol 2007;366:1209–1221.
3. Slovic AM, Kono H, Lear JD, Saven JG, DeGrado WF. Computational design of water-soluble analogues of the potassium channel KcsA. Proc Natl Acad Sci USA 2004;101:1828–1833.
4. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science 2003;302:1364–1368.
5. Watters AJ, Deka P, Corrent C, Callender D, Varani G, Sosnick T, Baker D. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. Cell 2007;128:613–624.
6. Dallüge R, Oschmann J, Birkenmeier O, Lücke C, Lilie H, Rudolph R, Lange C. A tetrapeptide fragment-based design method results in highly stable artificial proteins. Proteins 2007;68:839–849.
7. Pattabiraman N, Ward KB, Fleming PJ. Occluded molecular surface: analysis of protein packing. J Mol Recognit 1995;8:334–344.
8. Maestre-Martínez M, Edlich F, Jarczowski F, Weiwad M, Fischer G, Lücke C. Solution structure of the FK506-binding domain of human FKBP38. J Biomol NMR 2006;34:197–202.
9. Houben K, Dominguez C, Van Schaik FMA, Timmers HT, Bonvin AMJJ, Boelens R. Solution structure of the ubiquitin-conjugating enzyme UbcH5B. J Mol Biol 2004;344:513–526.
10. Cole R, Loria JP. FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. J Biomol NMR 2003;26:203–213.
11. Wüthrich K. NMR of proteins and nucleic acids. New York: Wiley; 1986.
12. Pristovšek P, Lücke C, Reincke B, Ludwig B, Rüterjans H. Solution structure of the functional domain of *Paracoccus denitrificans* cytochrome c_{552} in the reduced state. Eur J Biochem 2000;267:4205–4212.
13. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273:283–298.
14. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
15. Dantas G, Watters AL, Lunde BM, Eletr ZM, Isern NG, Roseman T, Lipfert J, Doniach S, Tompa M, Kuhlman B, Stoddard BL, Varani G, Baker D. Mis-translation of a computationally designed protein yields an exceptionally stable homodimer: implications for protein engineering and evolution. J Mol Biol 2006;362:1004–1024.
16. Richardson JS, Richardson DC. Amino acid preferences for specific locations at the ends of α helices. Science 1988;240:1648–1652.