

Biochemie 2004

Das rationale Proteindesign mit rechnergestützten Methoden ist in eine überraschend produktive Phase getreten: Nun können stabilisierte Proteine und sogar künstliche Enzyme experimentell realisiert werden. Röntgenstrukturanalysen machen die Erkennung von Nucleinsäuren durch Proteine in der DNA-Replikation und bei Prozessen der RNA-Interferenz im molekularen Detail sichtbar.

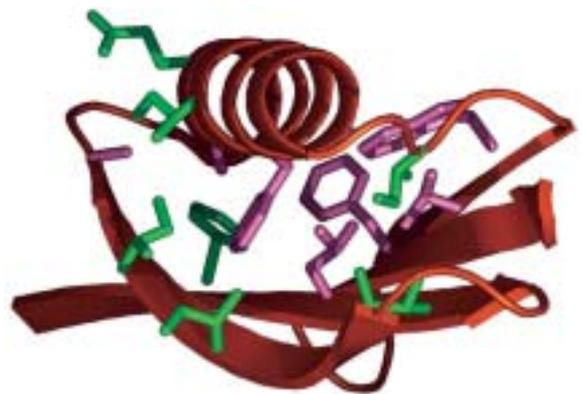
Rechnergestütztes Proteindesign

◆ Die enormen Fortschritte des rationalen Designs von Proteinen beginnen, unser Verständnis der Biologie zu revolutionieren. Das Wissen um die biochemischen Mechanismen auf atomarem Niveau, verbunden mit der rasanten Entwicklung von Algorithmen, gesteigerter Rechenleistung und verbesserter Parametrisierung von Kraftfeldern, erlauben es nun, künstliche Proteine mit neuartigen Funktionen zu entwerfen.¹⁾ Damit tritt ein Wandel von der Aufklärung und Beschreibung hin zum kreativen Gestalten von Proteinstrukturen ein. So wie die Manipulation des genetischen Codes die erste Revolution in den Biowissenschaften eingeleitet hat, so markiert auch dieser Umbruch den Beginn einer aufregenden Entwicklung.

Zwar liefert die astronomische Zahl der Kombinationsmöglichkeiten der 20 natürlichen Aminosäurebausteine in einer Polypeptidkette enorme Möglichkeiten für neuartige Wirkstoffe, doch resultiert daraus auch eine kaum lösliche Rechenaufgabe hinsichtlich der Vorhersage von Struktur und Funktion. Einfache Abschätzungen zeigen, daß selbst unter optimalen Bedingungen weniger als 10^{55} unterschiedliche Polypeptide auf der Erde entstanden sein

und in der Evolution auf Verwendbarkeit geprüft werden konnten.²⁾ Diese Zahl ist allein schon angesichts der etwa 10^{130} kombinatorisch möglichen Proteindomänen aus bloß 100 Aminosäuren ein verschwindend kleiner Ausschnitt. Unter den also etwa 10^{130} mangels Zeit und Ressourcen unbekannt gebliebenen Polypeptiden finden sich gewiß zahlreiche potentiell nützliche Proteine; die Frage ist nur, wie diese im abstrakten Sequenzraum mit theoretischen Methoden zu lokalisieren sind.

Lange blieb Proteindesign in erster Linie ein hehres Ziel, eng verbunden mit der universellen Lösung des Proteinfaltungsproblems. Diese Suche nach dem „Stein der Weisen“ ist inzwischen pragmatischen Lösungsansätzen gewichen. Mit rund 30 000 Strukturen in der Proteindatenbank und dem weltweiten Anlaufen der strukturellen Proteomforschungsprojekte ist die Datengrundlage so umfangreich geworden, daß sich die Zahl der in der Natur vertretenen Proteinarchitekturen relativ verlässlich auf etwa 2000 abschätzen läßt.³⁾ Zunehmend gehören statistisch basierte Methoden wie Threading und Fold-Recognition – neben der klassischen Energieminimierung und Molekulardynamik – zum unverzichtbaren Instrumentarium bei der Vorhersage und Planung von Proteinstrukturen.¹⁾



Seitenkettenpackung und Dead-End-Elimination-Theorem

◆ Das Konzept von der „inversen Proteinfaltung“, also der Suche nach Aminosäuresequenzen, die sich in eine vorgegebene Tertiärstruktur falten, hat dem rechnergestützten Proteindesign wichtige Impulse gegeben. Als nützliches Werkzeug erwies sich dabei das Dead-End-Elimination-Theorem (DEET).⁴⁾

Die Vorhersage von Seitenkettenkonformation in Proteinen ist von zentraler Bedeutung sowohl für das Design neuer Proteine als auch für das Homology-Modelling von natürlichen Proteinen anhand verwandter Biomoleküle mit bekannter Raumstruktur. In vereinfachter Form stellt sich diese Aufgabe als das „Packungsproblem“: Ausgehend von einer bekannten Konformation des Polypeptid-Rückgrats sollen Seiten-

Abb. 1. Packungsoptimierung mit DEET in der $\beta 1$ -Domäne von Protein G, dargestellt anhand der NMR-Struktur (PDB-Code 1GB4).⁸⁾ Die Seitenketten der sieben mutierten Reste in der thermostabilen Variante sind grün, die anderen Reste aus dem hydrophoben Kernbereich rosa wiedergegeben.

ketten in energetisch günstiger Geometrie eingesetzt werden. Trotz der durchaus diskreten bevorzugten Diederwinkelkombinationen in den einzelnen Aminosäureseitenketten – auch als „Rotamere“ bezeichnet – macht die unübersehbare Zahl möglicher Seitenkettenkombinationen eine Strukturvorhersage sehr schwierig.

Dem lange als kaum lösbar angesehenen Problem, hierbei das absolute Energieminimum berechnen zu können, wurde mit dem DEET auf wirkungsvolle Weise begegnet. Dabei werden iterativ paarweise Kombinationen von solchen Rotameren eliminiert, welche nicht zum globalen Energieminimum führen können. Dies reduziert den rechentechnischen Aufwand drastisch.^{4,5)} Mit neueren Algorithmen können kleine bis durchschnittlich große Protein-domänen in kurzer Zeit automatisch berechnet werden.⁶⁾ Anwendungen des DEET im Proteindesign haben sich zunächst auf die optimierte Packung der Seitenketten in der hydrophoben Kernregion von Proteinen (Core-Repacking) zum Zweck der verbesserten Faltungstabilität konzentriert.^{6,7)}

Mit dieser Strategie ließ sich unter anderem eine hitzestabile Variante aus dem mesophilen *Streptococcus*-Protein G (genauer: seiner β 1-Domäne mit 57 Resten) entwickeln, welche eine Schmelztemperatur von über 99 °C aufweist.⁸⁾ Dessen praktisch bedeutsame Bindungsaktivität gegenüber humanen Immunglobulinen blieb

dabei weitgehend erhalten. Die durch bloß sieben Aminosäuresubstitutionen erreichte hohe Hitze-stabilität war zurückzuführen auf eine optimierte Core-Packung, Abschirmung hydrophober Reste, günstigere Interaktionen des Helixdipols sowie verbesserte Tendenz zur Bildung von Sekundärstrukturen (Abbildung 1).

Von der Konstruktion von Bindungsstellen ...

◆ Die Komplexbildung aus Protein und Ligand, ob makro- oder niedermolekularer Natur, ist ein fundamentaler biologischer Prozess, der auf atomarem Niveau gesteuert wird. Manipulationen dieser Vorgänge, seit den frühesten Tagen des Protein-Engineerings praktiziert, erweitern nicht nur das Wissen über die zugrundeliegenden Mechanismen, sondern lassen auch an die Konstruktion von künstlichen Bindungsproteinen oder gar die Entwicklung von neuartigen Enzymen denken.

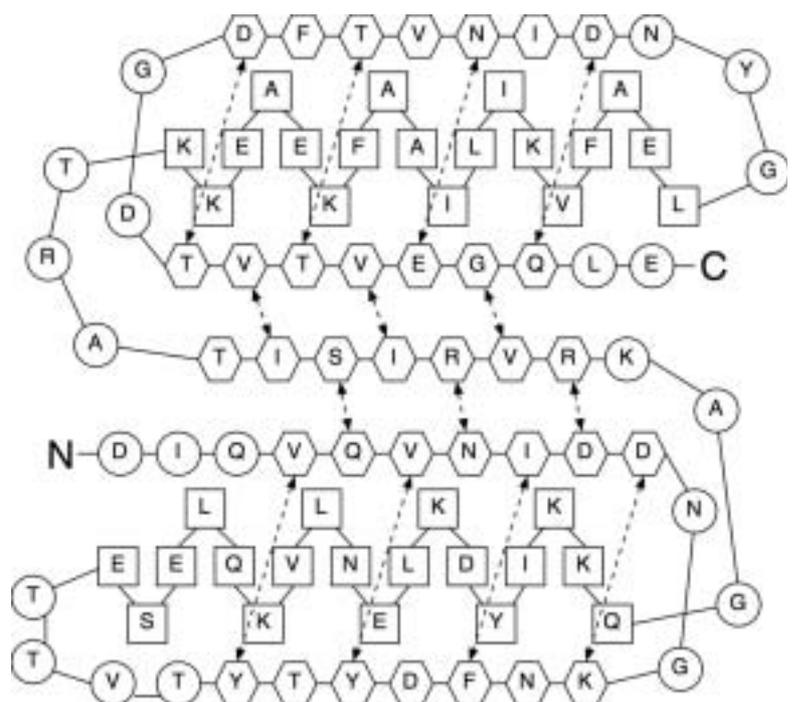
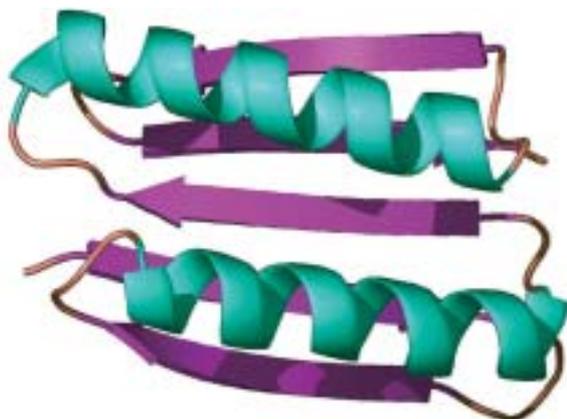
Die Berechnung der optimierten Seitenkettenpackung im Innern eines Proteins unter Anwendung des DEET kann prinzipiell auch auf die Anordnung von Aminosäuren an der

Grenzfläche zweier Proteine, welche einen nicht kovalenten Komplex miteinander eingehen, ausgedehnt werden. So gelang es im Fall des Calmodulins, die Spezifität zu einem seiner zahlreichen Liganden, der leichten Kette der Myosinkinase smMLCK, zu verbessern.⁹⁾ Die beste konstruierte Version wies eine 155-fache Steigerung der Bindungsspezifität auf, wobei fünf Aminosäuren von acht an der Kontaktbildung beteiligten Resten in der Bindungsstelle des Calmodulins ausgetauscht waren.

Eine kürzlich erschienene Arbeit befaßte sich mit der Umgestaltung der Bindungsstelle der bakteriellen DNase Colicin E7 für deren Inhibitor Im7.¹⁰⁾ Austausch von Seitenketten an beiden Proteinen führte zu einer neuen Bindungsspezifität, deren strukturelle Basis auch proteinkristallographisch analysiert wurde. Die neu generierte Spezifität war sogar in vivo nachweisbar, trotz der großen Zahl möglicher konkurrierender Proteininteraktionen in einer Zelle.

Die Anwendung von DEET-Algorithmen auf eine Familie untereinander homologer Metabolit-Bindungsproteine mit bekannter Raumstruktur aus dem Periplasma von

Abb. 2. Rationale Konstruktion eines α/β -Proteins mit 93 Aminosäuren.¹⁵⁾ Rechts: schematische Planung von „Top7“: Sechsecke kennzeichnen β -Faltblattstränge, Quadrate α -Helices und Kreise Schleifenregionen. Rückgrat-Wasserstoffbrücken zwischen zwei Aminosäuren sind als gestrichelte Doppelpfeile dargestellt. Links: Kristallstruktur des synthetischen Proteins (PDB-Code 1QYS).



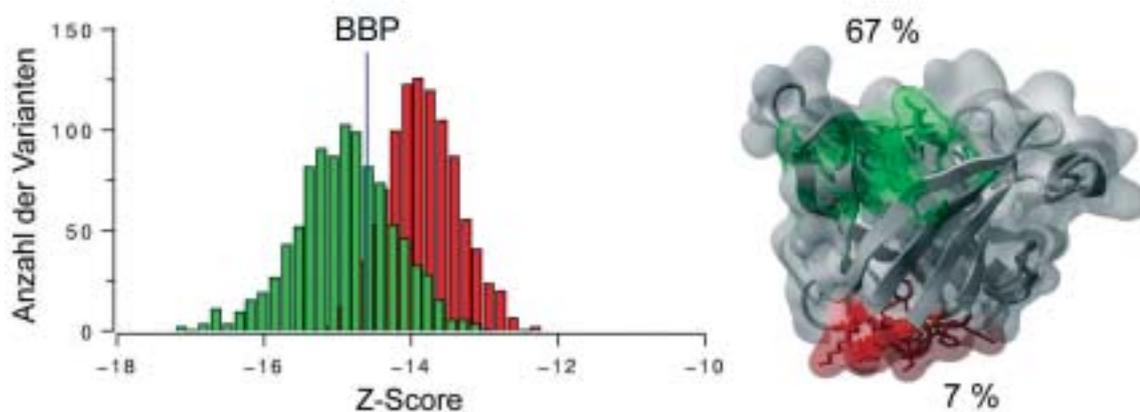


Abb. 3. Berechnete Toleranz eines Protein-Scaffolds gegenüber zufälligen Aminosäuresubstitutionen im Hinblick auf die Konstruktion von Bibliotheken zur Selektion künstlicher Bindungsproteine.¹⁷⁾ Als Beispiel gezeigt ist das Bilin-Bindungsprotein (BBP), welches bereits als Gerüststruktur zur Herstellung von „Anticalinen“ mit neuartigen Ligandenspezifitäten gedient hat.¹⁶⁾ In der Tertiärstruktur sind diejenigen 16 Reste grün eingefärbt, die zur experimentellen Generierung von spezifischen Bindungsstellen für vorgegebene niedermolekulare Liganden wie Fluorescein oder Digoxigenin ausgetauscht worden sind. Eine Kontrollregion mit ebenfalls 16 Resten, aber am geschlossenen Ende der β -Faltstruktur, ist rot markiert. Das Histogramm zeigt, daß bei simulierter zufälliger Substitution von Seitenketten im Bereich der Bindungsregion 67 % der Varianten gleich große oder sogar bessere Stabilität (entsprechend stärker negativem Z-Score) aufweisen. Im Fall der Kontrollregion sind dagegen 93 % der Varianten gegenüber dem unmutierten BBP destabilisiert. (Graphik mit freundlicher Genehmigung von Markus Wiederstein und Manfred J. Sippl, Universität Salzburg).

Escherichia coli hat es ermöglicht, deren Bindungseigenschaften für niedermolekulare Liganden zu variieren.¹¹⁾ Nach Neupackung von 12 bis 18 Aminosäuren im Bereich der Bindungstasche im Hinblick auf strukturelle Komplementarität und Wasserstoffbrücken gegenüber den neuen Liganden, die ähnliche Raumaufüllung wie die natürlichen Liganden (z. B. Ribose) hatten, konnten Rezeptorproteine hergestellt werden, welche Trinitrotoluol, L-Lactat oder Serotonin mit hoher Selektivität und Affinität um 1 μ M komplexieren. Da die periplasmatischen Bindungsproteine auf die Komplexierung ihres jeweiligen Liganden mit einer globalen Konformationsänderung reagieren, ließ sich die ortsspezifische Kopplung mit einem Styrol-Fluoreszenzfarbstoff zudem als Indikator für den Bindungszustand des Liganden nutzen. Dies ermöglichte die Entwicklung einfacher Biosensoren.¹²⁾

... zum Design künstlicher Enzyme

◆ Weitere Designarbeiten an den periplasmatischen Bindungsproteinen gipfelten in der Konstruktion ei-

ner biologisch aktiven Triosephosphat-Isomerase auf der Grundlage des Ribosebindungsproteins, welches selbst keine Enzymaktivität aufweist.¹³⁾ Zunächst wurde dessen Ligandenbindungsstelle mit den zuvor entwickelten theoretischen Methoden im Hinblick auf die Komplexierung sowohl von Glycerinaldehydphosphat als auch von Dihydroxyacetonphosphat umgestaltet. Damit wurden mikromolare Dissoziationskonstanten erreicht. Danach wurden katalytisch aktive Reste (Glutamat, Histidin, Lysin), wie sie aus der völlig anders gefalteten natürlichen Triosephosphat-Isomerase bekannt sind, in geeigneter Geometrie eingeführt.

Eines der insgesamt 14 experimentell untersuchten Proteine, NovoTim1.0, zeigte tatsächlich schwache Enzymaktivität. Da NovoTim1.0 eine erheblich verminderte Faltungstabilität aufwies ($T_m = 37^\circ\text{C}$), konzentrierte sich die weitere Optimierung auf die Stabilisierung des artifizialen Enzyms mittels Packungsoptimierung, teilweise fernab des aktiven Zentrums, wobei sich die enzymatischen Eigenschaften überraschenderweise verbesserten.

Durch experimentelle Zufallsmutagenese und Selektion in vivo konnte schließlich eine Mutante identifiziert werden, die das Wachstum eines TIM-defizienten Stamms von *Escherichia coli* auf einem Glycerin-Minimalmedium unter Bedingungen der Gluconeogenese rekonstituierte.

Diese erfolgreiche Konstruktion eines Enzyms mit einer typischen biochemischen Aktivität nach hauptsächlich rationalen Prinzipien dürfte wegweisend für das künftige Design von Biokatalysatoren sein, vor allem, nachdem frühere Versuche zur Konstruktion von neuen Enzymen allein mit der kombinatorischen Biochemie weitgehend fruchtlos geblieben sind.¹⁴⁾

Neben der Generierung katalytischer Aktivität war der Entwurf von Proteinen mit neuartiger Raumstruktur lange die größte Herausforderung für das computergestützte Proteindesign. Mit der Konstruktion eines aus 93 Aminosäuren aufgebauten künstlichen Proteins namens Top7, das eine bislang unbekannte α/β -Topologie aufweist (Abbildung 2), ist dieses Ziel wohl zum ersten Mal erreicht worden.¹⁵⁾ Das aus einem Prozeß iterativer Struk-

vorhersage und Sequenzanpassung, ausgehend von einem willkürlich vorgegebenen groben Strukturmodell, resultierende Protein zeigt das erwartete CD-Spektrum und ist gegenüber thermischer sowie chemischer Entfaltung überraschend stabil. Auch die experimentell aufgeklärte Kristallstruktur ist dem vorhergesagten Modell sehr ähnlich.

Damit ist nachgewiesen, daß es mit rationalem Proteindesign tatsächlich möglich ist, nicht nur natürlich vorkommende, sondern auch artifizielle Faltungstypen zu erzeugen und zu optimieren. Auf diese Weise dürfte sich letztlich der große noch unerforschte Proteinsequenzraum einer praktischen Anwendung erschließen lassen.

Die Zukunft des Proteindesigns

◆ Nach den Erfolgen der experimentellen Methoden der kombinatorischen Biochemie, manchmal auch despektierlich „irrationales“ Proteindesign genannt, haben die rechnergestützten rationalen Methoden in den letzten Jahren hinsichtlich ihrer Einsatztauglichkeit erheblich an Bedeutung gewonnen. Der Vorteil des kombinatorischen Proteindesigns beruhte im wesentlichen auf den Möglichkeiten der Gentechnik, umfangreiche – aber durchaus zahlenmäßig limitierte – Repertoires an Mutanten eines Proteins parallel zu erzeugen und da-

raus Varianten mit verbesserten oder gar neuen Eigenschaften zu selektieren. Dieses Verfahren wird heute vor allem bei der Konstruktion von künstlichen Bindungsproteinen angewandt.¹⁶⁾

Effiziente numerische Verfahren, vor allem die DEET-Algorithmen zur Seitenkettenpackung sowie die statistisch basierten Fold-Recognition- und Mean-Force-Field-Methoden, sind inzwischen soweit gereift, daß sie sogar eine Auswahl derjenigen Sequenzkombinationen gestatten, die durch kombinatorische Biotechnologie sinnvollerweise experimentell analysierbar sind. Damit sollte sich die Trefferwahrscheinlichkeit in einem molekularbiologischen Selektionsexperiment wesentlich steigern lassen.

In diesem Zusammenhang hat eine rechnergestützte Analyse jüngst gezeigt, daß es keinesfalls bloß von offensichtlichen strukturellen Kriterien wie dem Vorhandensein variabler Schleifenregionen oder exponierter Aminosäureseitenketten abhängt, ob sich Proteinarchitekturen als „Scaffolds“ (Gerüststrukturen) zur Generierung neuer aktiver Zentren eignen.¹⁷⁾ Vielmehr ist die Toleranz für konstruktive Eingriffe offenbar in der Faltung des Proteins selbst angelegt, was sich nicht nur für die Antigenbindungsstellen der Antikörper, sondern beispielweise auch für die Ligandenbindungsstellen der Lipocalin-Proteine eindrucksvoll belegen läßt (Abbildung 3).

Es ist abzusehen, daß rechnergestützte Methoden beim Proteindesign ihre ursprünglich erhoffte Rolle bald einnehmen werden. Mit einer rasant wachsenden Basis an Strukturdaten sowie kontinuierlich verfeinerten theoretischen Methoden läßt sich der Zeitpunkt erahnen, an dem „Computereperimenten“ gegenüber der Laborarbeit der Vortzug zu geben ist.

Arne Skerra, Andreas Eichinger
Lehrstuhl für Biologische Chemie
Technische Universität München
Freising-Weihenstephan
skerra@wzw.tum.de
a.eichinger@wzw.tum.de



Andreas Eichinger (Jg. 1965) studierte Biologie in Regensburg und Amsterdam. Er schloß seine Doktorarbeit im Jahre 1999 in der Arbeitsgruppe Strukturfor-

schung am Max-Planck-Institut für Biochemie in Martinsried ab. Danach ging er für einen Forschungsaufenthalt zuerst an das Burnham Institute und anschließend an das Salk Institute in La Jolla, Kalifornien. Seit 2003 arbeitet er auf dem Gebiet der Proteinkristallographie am Lehrstuhl für Biologische Chemie der TU München.



Arne Skerra (Jg. 1961) hat in Darmstadt und München Chemie studiert. Nach seiner Promotion am Gen-Zentrum der Ludwig-Maximilians-Universität

arbeitete er am Laboratory of Molecular Biology in Cambridge, England. Anschließend baute er eine Arbeitsgruppe am MPI für Biophysik in Frankfurt/M. auf. 1994 erhielt er einen Ruf an die TU Darmstadt, wo er die Abt. Proteinchemie leitete. Seit 1998 ist er Ordinarius an der TU München. 2003 wurde er in den Vorstand der Fachgruppe Biochemie der GDCh gewählt.

- 1) N. Pokala, T. M. Handel, *J. Struct. Biol.* 2001, 134, 269.
- 2) H. M. Wilks, A. Cortes, D. C. Emery, D. J. Halsall, A. R. Clarke, J. J. Holbrook, *Ann. NY Acad. Sci.* 1992, 672, 80.
- 3) S. Govindarajan, R. Recabarren, R. A. Goldstein, *Proteins* 1999, 35, 408.
- 4) J. Desmet, M. De Maeyer, B. Hazes, I. Lasters, *Nature* 1992, 356, 539.
- 5) I. Lasters, M. De Maeyer, J. Desmet, *Protein Eng.* 1995, 8, 815.
- 6) L. L. Looger, H. W. Hellinga, *J. Mol. Biol.* 2001, 307, 429.
- 7) S. Ventura, L. Serrano, *Proteins* 2004, 56, 1.
- 8) S. M. Malakauskas, S. L. Mayo, *Nature Struct. Mol. Biol.* 1998, 5, 470.
- 9) J. M. Shifman, S. L. Mayo, *Proc. Natl. Acad. Sci. USA* 2003, 100, 13274.
- 10) T. Kortemme, L. A. Joachimiak, A. N. Bullock, A. D. Schuler, B. L. Stoddard, D. Baker, *Nature Struct. Mol. Biol.* 2004, 11, 371.
- 11) L. L. Looger, M. A. Dwyer, J. J. Smith, H. W. Hellinga, *Nature* 2003, 423, 185.
- 12) R. M. de Lorimier, J. J. Smith, M. A. Dwyer, L. L. Looger, K. M. Sali, C. D. Paavola, S. S. Rizk, S. Sadigov, D. W. Conrad, L. Loew, H. W. Hellinga, *Protein Sci.* 2002, 11, 2655.
- 13) M. A. Dwyer, L. L. Looger, H. W. Hellinga, *Science* 2004, 304, 1967.
- 14) M. M. Altamirano, J. M. Blackburn, C. Aguayo, A. R. Fersht, *Nature* 2002, 417, 468.
- 15) B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, D. Baker, *Science* 2003, 302, 1364.
- 16) A. Skerra, *Curr. Opin. Chem. Biol.* 2003, 7, 683.
- 17) M. Wiederstein, M. J. Sippl, *J. Mol. Biol.* 2005, 345, 1199.