

Analoga ist damit ebenso zuverlässig machbar wie die Synthese in größerem Maßstab.

Verwendete Abkürzungen

Abu: 2-Aminobuttersäure, Boc: *tert*-Butoxycarbonyl, BTC: Bis(trichlormethyl)carbonat, DCC: *N,N'*-Dicyclohexylcarbodiimid, DIC: *N,N'*-Diisopropylcarbodiimid, EDC: *N*-Ethyl-*N'*-(3-dimethylaminopropyl)carbodiimid-hydrochlorid, Fmoc: 9-Fluorenylmethoxycarbonyl, HATU: O-(7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium-hexafluorophosphat (IUPAC: 1-[Bis-(dimethylamino)methylumyl]-1*H*-1,2,3-triazolo[4,5-*b*]pyridin-3-oxid-hexafluorophosphat), HOAt: 1-Hydroxy-7-azabenzotriazol, MeBmt: (4*R*)-4[(*E*)-2-

Butenyl-4,*N*-dimethyl-L-threonin, Nva: Norvalin, TFFH: Tetramethylfluorformamidinium-hexafluorophosphat, Trt: Triphenylmethyl (Trityl).

- [1] B. Thern, J. Rudolph, G. Jung, *Tetrahedron Lett.* **2002**, 43, 5013–5016.
- [2] B. Thern, J. Rudolph, G. Jung, *Angew. Chem.* **2002**, 114, 2401–2403; *Angew. Chem. Int. Ed.* **2002**, 41, 2307–2309.
- [3] N. Sewald, H.-D. Jakubke, *Peptides: Chemistry and Biology*, Wiley-VCH, Weinheim, **2002**.
- [4] E. Falb, Y. Yechezkel, Y. Salitra, C. Gilon, *J. Pept. Res.* **1999**, 53, 507–517.
- [5] H. Eckert, B. Forster, *Angew. Chem.* **1987**, 99, 922–923; *Angew. Chem. Int. Ed. Engl.* **1987**, 26, 894–895.
- [6] L. A. Carpino, M. Beyermann, H. Wenschuh, M. Bienert, *Acc. Chem. Res.* **1996**, 29, 268–274.
- [7] L. A. Carpino, A. El-Faham, *Tetrahedron* **1999**, 55, 6813–6830.

Neue Proteinstrukturprinzipien: Nester, Eier – und was noch?*

Debnath Pal, Jürgen Sühnel* und Manfred S. Weiss*

Rückfaltungsexperimente an durch Harnstoff denaturierter Ribonuclease, die vor mehr als 40 Jahren durchgeführt wurden, veranlassten den Chemie-Nobelpreisträger Anfinsen zu folgender Aussage, die das noch heute weitgehend gültige Paradigma der Proteinfaltung widerspiegelt: „... *it may be concluded that the information ... for the assumption of the native secondary and tertiary structures* [von Proteinen] *is contained in the amino acid sequence itself*.“^[1] Davon ausgehend sollte man eigentlich in der Lage sein, die dreidimensionale Struktur von Proteinen direkt aus ihrer Sequenz vorherzusagen. Allerdings sind trotz intensiver Bemühungen vieler ausgezeichneten Wissenschaftler und einer Datenbasis experimentell bestimmter Proteinstrukturen, die in geradezu beängstigender Weise täglich größer wird,^[2] wirkliche Erfolge bei der Vorhersage von Proteinstrukturen selten. Die gegenwärtige Situation stellt sich daher als eher enttäuschend dar. Die Gründe dafür sind unklar. Trotz der schon erwähnten umfangreichen experimentellen Datenbasis, die durch die

rasante Entwicklung der Strukturbiochemie in den letzten zehn Jahren aufgebaut wurde, und trotz vieler gründlicher Analysen dieser Daten^[3–5] ist unsere gegenwärtige Kenntnis von Proteinstrukturen nach wie vor in erster Linie deskriptiver Natur und ohne ein wirkliches Vorhersagepotenzial. Eine mögliche Erklärung dafür könnte sein, dass die gegenwärtig bekannten Proteinstrukturprinzipien unvollständig sind. Somit bestünde ein Bedarf an neuen Konzepten, um von dem heutigen vorwiegend deskriptiven Status zu wirklichen Vorhersagen zu kommen.

In dieser Hinsicht erscheinen zwei Arbeiten, die kürzlich im *Journal of Molecular Biology* publiziert wurden, von besonderem Interesse.^[6, 7] Bei einer Analyse der Hauptkettentorsionswinkel benachbarter Aminosäuren fanden Watson und Milner-White, dass viele Bindungsstellen für Anionen und Kationen aus drei in der Sequenz aufeinander folgenden Aminosäuren aufgebaut sind (wobei unter Anionen und Kationen alle Atome zu verstehen sind, die entweder eine volle oder eine partielle negative bzw. positive Ladung tragen). Zwei benachbarte dieser drei Aminosäuren haben dabei „enantiomere“ Hauptkettenkonformationen. Der Terminus „enantiomer“ wurde gewählt, weil die Werte der Hauptkettentorsionswinkel (ϕ, ψ) der beiden Aminosäuren bezüglich des Koordinatenursprungs des Ramachandran-Plots näherungsweise invertiert sind.^[3] Während in der Sequenz aufeinander folgende Aminosäuren mit identischen oder nahezu identischen Hauptkettenkonformationen α -Helices, β -Stränge oder Polyprolin-Typ-II-Helices bilden, ergeben benachbarte Aminosäuren mit enantiomeren Hauptkettenkonformationen so genannte „Nester“ oder Nest-Motive. Die Hauptkettentorsionswinkel (ϕ, ψ) der entsprechenden

[*] Dr. J. Sühnel, Dr. D. Pal
Institut für Molekulare Biotechnologie
Beutenbergstraße 11, 07745 Jena (Deutschland)
Fax: (+49) 3641-656210
E-mail: jsuehnel@imb-jena.de

Dr. M. S. Weiss
EMBL Hamburg Outstation
c/o DESY, Notkestraße 85, 22603 Hamburg (Deutschland)
Fax: (+49) 40-89902-149
E-mail: msweiss@embl-hamburg.de

[**] Die Autoren bedanken sich bei E. James Milner-White für die Einführung in das Nest-Konzept und für weitere stimulierende Diskussionen.

Aminosäurenachbarn nehmen dabei Werte von etwa $(-90^\circ, 0^\circ)$ und $(+90^\circ, 0^\circ)$ oder umgekehrt an.

Der Begriff „Nest“ ist davon abgeleitet, dass die NH-Gruppen der drei aufeinander folgenden Aminosäuren, von denen zwei das oben genannte Torsionswinkelkriterium befolgen, eine Vertiefung bilden, die als Bindungsstelle für ein Atom oder eine Gruppe von Atomen mit einer ganzen oder partiellen negativen Ladung dienen kann. Je nachdem, welche Kombination von Torsionswinkelpaaren beobachtet wird, unterscheidet man RL-Nester ($\phi_1, \psi_1 = -90^\circ, 0^\circ$; $\phi_2, \psi_2 = +90^\circ, 0^\circ$) und LR-Nester ($\phi_1, \psi_1 = +90^\circ, 0^\circ$; $\phi_2, \psi_2 = -90^\circ, 0^\circ$). Zwei oder auch mehrere dieser Motive können miteinander kombiniert sein und auf diese Weise zusammengesetzte Nester (compound nests) oder aufeinander folgende Nester (tandem nests) mit insgesamt bis zu acht Aminosäuren bilden. In der Mehrzahl aller Fälle binden die Nest-Motive ein Atom oder eine Gruppe von Atomen, die wir in Analogie zu dem Begriff „Nest“ als „Ei“ bezeichnen. Es ist ausgesprochen interessant, dass viele bekannte Struktur motive wie Schellman-Schleifen, die Oxyanionenbindungstasche in Serin-Proteasen oder auch P-Schleifen in ATP- oder GTP-bindenden Proteinen in das Nest/Ei-Konzept eingeordnet werden können. Berücksichtigt man weitere enantiomere Kombinationen von Hauptkettentorsionswinkeln in Dipeptiden, so kann dieses Konzept auch auf Kationenbindungsstellen ausgedehnt werden.^[7] Allerdings gibt es dafür bisher viel weniger experimentelle Beispiele als für Anionenbindungsstellen.^[6]

Watson und Milner-White^[6, 7] haben ihre Analyse mit einem relativ kleinen Datensatz von 67 Proteinstrukturen durchgeführt. Eine erweiterte Analyse ihres Konzepts mit einem etwa zwanzigfach größeren Datensatz, der von uns bereits in einer anderen Arbeit^[8] verwendet wurde, bestätigt die meisten Schlussfolgerungen dieser Autoren, fügt aber auch einige neue Aspekte hinzu. Wie oben bereits erwähnt, können Nest-Motive durch die Hauptkettentorsionswinkel zweier aufeinander folgender Aminosäuren beschrieben werden. Eine aussagekräftigere Möglichkeit ergibt

sich aber aus der Betrachtung der Torsionswinkel zwischen den N-H-Gruppen von jeweils zwei der Nest-Aminosäuren und des Winkels zwischen den drei N-Atomen. Dabei wird zusätzlich eine flankierende Aminosäure berücksichtigt (Abbildung 1). Im Unterschied zum Ramachandran-Plot, in dem die R- und L-Regionen der RL- und LR-Nester naturgemäß überlappen, erscheinen die beiden Nest-Klassen in der in Abbildung 1 gezeigten dreidimensionalen Darstellung als deutlich voneinander getrennte Inseln. In Tabelle 1 sind die Mittelwerte der geometrischen Parameter aufgeführt. Zwei

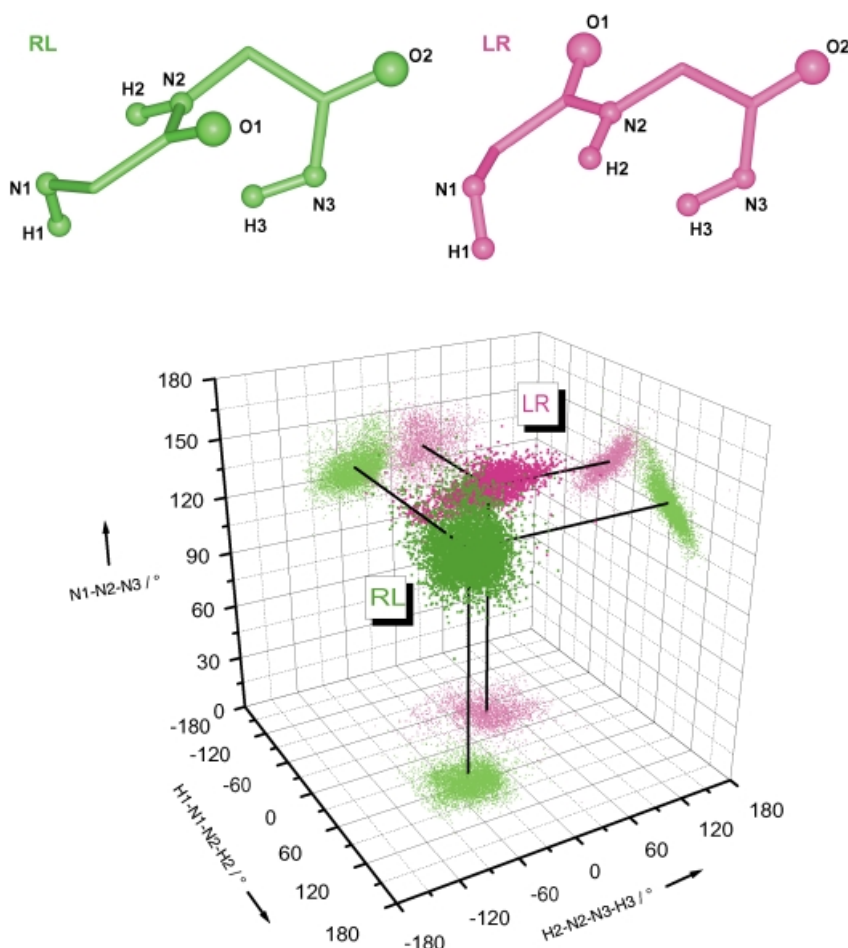


Abbildung 1. Dreidimensionale Darstellung aller Aminosäurenpaare, die das Nest-Kriterium erfüllen. Die drei Achsen sind: a) Torsionswinkel, definiert durch die vier Atome H1-N1-N2-H2, b) Torsionswinkel, definiert durch H2-N2-N3-H3, c) Winkel zwischen den drei Atomen N1-N2-N3. Die RL-Nest-Motive sind in Grün und die LR-Nest-Motive in Rosa dargestellt. Die jeweiligen Projektionen werden in etwas hellerer Färbung gezeigt. Die Mittelwerte der oben genannten Parameter sind graphisch durch schwarze durchgehende Linien verdeutlicht, die im Zentrum der Verteilungen beginnen und auf die drei Flächen des Koordinatensystems projiziert werden. Außerdem sind zwei repräsentative Strukturen gezeigt, die etwa den Mittelwerten der geometrischen Parameter für das RL- und das LR-Motiv entsprechen.

Tabelle 1. Häufigkeit und mittlere geometrische Parameter der Nest-Motive.

	Häufigkeit ^[a]	ϕ_1	ψ_1	ϕ_2	ψ_2	$\angle \text{H}_1\text{N}_1\text{N}_2\text{H}_2$	$\angle \text{H}_2\text{N}_2\text{N}_3\text{H}_3$	$\angle \text{N}_1\text{N}_2\text{N}_3$
RL ^[b]	5773	-91°	-5°	$+74^\circ$	$+23^\circ$	$+56^\circ$	-41°	123°
LR ^[b]	2492	$+72^\circ$	$+21^\circ$	-85°	-22°	-44°	$+38^\circ$	128°

[a] Der verwendete Datensatz enthielt 1154 Proteinketten, 285794 Aminosäuren und 280563 Dipeptide (terminale Aminosäuren sind in der Liste der Dipeptide nicht enthalten, weil nur ein Hauptkettentorsionswinkel für diese Reste definiert werden kann).^[8] [b] R: $-140^\circ < \phi < -20^\circ$; $-90^\circ < \psi < +40^\circ$; L: $+20^\circ < \phi < +140^\circ$; $-40^\circ < \psi < +90^\circ$; die Peptidbindung zwischen den beiden Aminosäuren der RL- und LR-Motive liegt in allen Fällen in der *trans*-Konformation vor. Die Mittelwerte der Winkel ϕ und ψ für die R- und die L-Region sind trotz des hier verwendeten größeren Datensatzes den von Watson und Milner-White^[6] angegebenen Werten sehr ähnlich.

repräsentative Strukturmodelle für jeweils ein RL- und ein LR-Nest sind in Abbildung 1 gezeigt. Diese lassen deutlich die konkave Anordnung der NH-Gruppen erkennen. Zwei Beispiele für Nest-Motive aus einer realen Proteinstruktur sind in Abbildung 2 dargestellt.

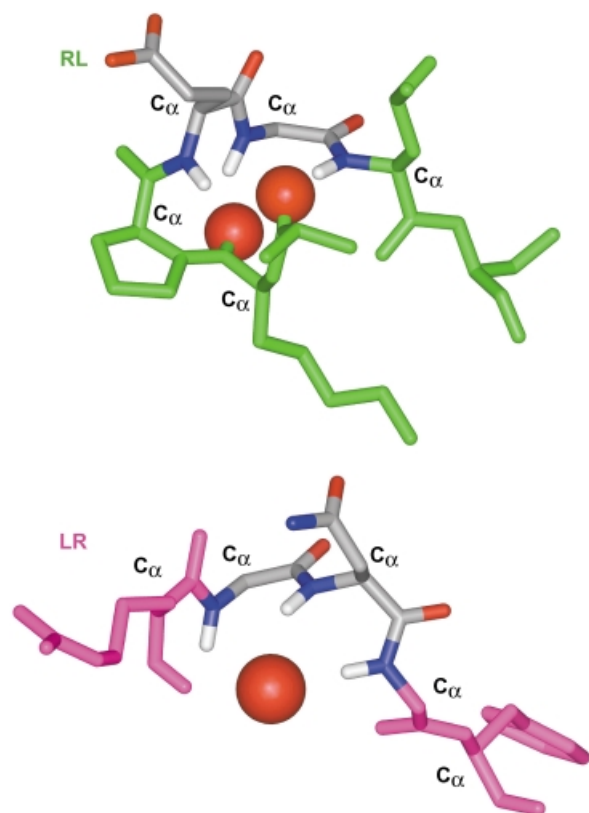


Abbildung 2. Beispiele für RL- und LR-Nest/Ei-Motive in einer Lysozym-Struktur (PDB-Code: 153L).^[10] Es werden jeweils die beiden Nest-Aminosäuren und die NH-Gruppe der dritten Aminosäure in atomspezifischer Färbung und die flankierenden Aminosäuren in Grün (RL) und Rosa (LR) gezeigt. Die Ei-O-Atome sind als rote Kugeln dargestellt. Das RL-Nest besteht aus den Aminosäuren Glu24, Gly25 und Leu27, und die entsprechenden Ei-Atome sind in diesem Fall die Carbonyl-O-Atome von Ala21 und Lys22. Das LR-Nest besteht aus Gly88, Asn89 und Gly90, und das entsprechende Ei-Atom ist ein Wasser-O-Atom (Wasser 216).

In Tabelle 1 sind neben den Mittelwerten der geometrischen Parameter auch die Häufigkeiten des Auftretens von Nest-Motiven in Proteinstrukturen angegeben. Insgesamt bilden 2.1% aller Dipeptide im Datensatz RL-Nester und 1.0% LR-Nester. Bezogen auf die Gesamtzahl von Aminosäuren im Datensatz sind insgesamt 5.4% an Nest-Motiven beteiligt. Die nächste überraschende Beobachtung ist, dass mehr als doppelt so viele Motive vom RL-Typ wie vom LR-Typ vorkommen. Das häufigere Vorkommen des RL-Motivs wurde bereits von Watson und Milner-White^[6] erwähnt; allerdings gaben sie ein größeres Verhältnis von 4:1 an, was wahrscheinlich durch den zu kleinen Datensatz bedingt war. Wir haben bisher keine einleuchtende Erklärung für die Bevorzugung von RL gegenüber LR. Immerhin scheint dies ein Hinweis darauf zu sein, dass die Hauptkettenkonformation einer Aminosäure in einem gewissen Ausmaß von der

Identität der Nachbarreste und von der Sequenzrichtung abhängt.

Da für alle Aminosäuren außer Glycin die L-Konformation energetisch ungünstig ist, gehört die Mehrheit (61%) der RL-Nester zum Typ Xaa-Gly und die Mehrheit (55%) der LR-Nester zum Typ Gly-Xaa. Insgesamt sind 21% aller Glycinreste in Proteinen an RL- oder LR-Nestern beteiligt, was eine mögliche Erklärung dafür sein könnte, dass Glycinreste oftmals besser konserviert sind als andere Aminosäuren in Proteinen.

Nest-Motive kommen bevorzugt an den Enden von α -Helices und inversen Schleifen (reverse turns) vor. Aus dem Torsionswinkelkriterium ergibt sich, dass der RL-Typ bevorzugt an den C-Termini von α -Helices angesiedelt ist, während der LR-Typ häufiger am N-Terminus vorkommt. Für Schleifen ist eine derartige Bevorzugung nicht zu beobachten. Wir haben außerdem nicht einen einzigen Fall gefunden, in dem ein Nest-Motiv an Schleifen vom Typ II oder II' beteiligt wäre.^[9]

Ein weitere wichtige Frage ist, wie viele Nest-Motive tatsächlich ein Ei enthalten, d. h. als Ionenbindungsstelle oder nach Watson und Milner-White^[6] als funktionale Nest-Motive fungieren. In unserer Analyse zeigte sich, dass 77% aller RL- und 52% aller LR-Nester mindestens ein Ligandenatom in einem Abstand enthalten, der einem H-Brücken-Abstand entspricht (Tabelle 2). In den meisten Fällen ist das Ei-Atom

Tabelle 2. Kategorien von Nest/Ei-Strukturen.

	RL	LR
Nest-Motive:		
Gesamtzahl	5773 (100%)	2492 (100%)
Besetzt	4419 (76.5%)	1290 (51.8%)
Nicht besetzt	1354 (23.5%)	1202 (48.2%)
Ei-Atome: ^[a]		
Gesamtzahl ^[b]	4838 (100%)	1386 (100%)
Hauptketten-Carbonyl-O-Atom	3009 (62.2%)	560 (40.4%)
Seitenketten-Carboxylat-O-Atom	630 (13.0%)	234 (16.9%)
Seitenketten-Amid-O-Atom	200 (4.1%)	44 (3.2%)
Seitenketten-Hydroxy-O-Atom	795 (16.4%)	49 (3.5%)
Ionen ^[c]	48 (1.0%)	55 (4.0%)
Wassermoleküle	156 (3.2%)	444 (32.0%)

[a] Nach Watson und Milner-White^[6] werden alle Liganden mit einem Abstand von weniger als 3.8 Å sowohl zu N1 als auch zu N3 berücksichtigt.
 [b] Die Gesamtzahl der Ei-Atome ist größer als die Zahl der besetzten Nest-Motive, weil manche Nest-Motive mehr als ein Ei enthalten.
 [c] Mögliche Ionen waren u. a.: Phosphat, Sulfat, Chlorid, Eisen-Schwefel-Cluster, phosphathaltige Cofaktoren. Beispiele sind in Tabelle 5 und Abbildung 5 in Lit. [6] zu finden.

ein Carbonyl-O-Atom von anderen Aminosäuren. Für Ei-Atome aus Seitenketten ergibt sich allerdings eine andere Situation. Bei Ser-OG und Thr-OG1 stammt mehr als die Hälfte der Ei-Atome aus den Nest-Aminosäuren selbst. Außer Atomen aus Aminosäuren werden in vielen Fällen solche aus Wassermolekülen und Ionen als Ei-Atome gefunden. Interessant ist dabei, dass nur 3% aller RL-Nester, aber 18% aller LR-Nester Wasser binden (Tabelle 2). Diese Befunde sprechen für eine allgemeine Rolle von Nest-Motiven als Bindungsstellen in Proteinen. Wahrscheinlich ist dabei die Kombination von Nest und Ei und nicht nur der Proteinteil allein als Strukturmotiv zu betrachten.

Zusammenfassend lässt sich feststellen, dass das von Watson und Milner-White entdeckte überraschend einfache Nest-Konzept einen wichtigen neuen Beitrag zu den Grundprinzipien der Architektur von Proteinen darstellt. Es fasst nicht nur eine Reihe bekannter Motive unter einem gemeinsamen Dach zusammen, sondern hat auch zur Identifizierung einiger neuer Motive geführt. Das Konzept ist nicht auf den reinen Proteinteil beschränkt, sondern berücksichtigt außerdem Liganden, Cofaktoren, Wassermoleküle usw. und liefert deshalb die Grundlage für eine generalisierte Betrachtung von Bindungsstellen in Proteinen. Man könnte sogar spekulieren, ob Nest-Motive frühe stabile Strukturen entlang des Faltungsweges von Proteinen darstellen. Da es sich um lokale Motive handelt, wäre der bei der Bildung auftretende Entropieverlust nur gering. Es bleibt zu hoffen, dass Konzepte wie dieses und möglicherweise weitere neue Ansätze unser derzeitiges Wissen über Proteinstrukturen erweitern und einen entscheidenden Beitrag leisten, um der derzeit noch

vorwiegend deskriptiven Strukturforschung an Proteinen eine stärker prädiktive Komponente zu verleihen.

-
- [1] C. B. Anfinsen, E. Haber, M. Sela, F. H. White, Jr., *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1309–1314.
 - [2] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.
 - [3] G. N. Ramachandran, V. Sasisekharan, *Adv. Protein Chem.* **1968**, *23*, 293–437.
 - [4] G. E. Schulz, R. H. Schirmer, *Principles of Protein Structure*, Springer, New York, **1979**.
 - [5] P. Chakrabarti, D. Pal, *Prog. Biophys. Mol. Biol.* **2001**, *76*, 1–102.
 - [6] J. D. Watson, E. J. Milner-White, *J. Mol. Biol.* **2002**, *315*, 187–198.
 - [7] J. D. Watson, E. J. Milner-White, *J. Mol. Biol.* **2002**, *315*, 199–207.
 - [8] M. Brandl, M. S. Weiss, A. Jabs, J. Sühnel, R. Hilgenfeld, *J. Mol. Biol.* **2001**, *307*, 357–377.
 - [9] T. E. Creighton, *Proteins: Structure and Molecular Properties*, W. H. Freeman, New York, **1993**.
 - [10] L. H. Weaver, M. G. Grütter, B. W. Matthews, *J. Mol. Biol.* **1995**, *245*, 54–68.
-