Modern total synthesis of protein molecules is based on the novel chemical ligation methods introduced by the Kent Laboratory at The Scripps Research Institute in the mid-1990s. Unprotected synthetic peptide segments spanning the amino acid sequence of the target polypeptide chain are covalently joined to one another in quantitative yield, without enzymes, by chemoselective reaction of unique, mutually reactive functional groups on each segment. Native chemical ligation – thioester-mediated amide-forming reaction at Cys residues – is the most robust and useful ligation chemistry developed to date. Chemical protein synthesis is straightforward and the outcome quite predictable; the challenge for most laboratories is making the peptide-thioester building blocks. We will discuss the high efficiency synthesis of a series of transmembrane peptide-thioesters, spanning the sequence of the protein diacylglycerol kinase, an integral membrane enzyme. We will also describe our approach to the total chemical synthesis of a GPCR protein and present preliminary results on the preparation of a set of peptide thioesters comprising the sequence of the δ-opioid receptor (δ-OR). Recent innovations that extend the range of targets to which chemical protein synthesis can be applied will be described. A series of case studies from our current work will be presented to illustrate the current capabilities of chemical protein synthesis and some of its applications. Chemical protein synthesis is uniquely enabling for the almost unlimited range of noncoded structures that can be introduced into the protein molecule, and for the site-specific labeling of proteins that enhances the application of advanced biophysical methods. We will illustrate these capabilities by the use of single molecule fluorescence spectroscopy and Ramachandran space engineering to further elucidate the mechanism of enzyme catalysis in the HIV-1 protease.

References
Total Chemical Synthesis of Proteins for Biological Research

Stephen Kent
Institute for Biophysical Dynamics
Department of Biochemistry & Molecular Biology
Department of Chemistry

University of Chicago

Roadmap Meeting
TSRI, La Jolla
November 2007
CHEMICAL PROTEIN SYNTHESIS

Sequence Data

Bioinformatics

Chemistry

Candidate Protein Sequences

Synthetic Polypeptide

Synthetic Protein

Biological Activity

Neutrophil migration
- NSA
- MetOF-1α

Concentration (nM)

Cl
Total Synthesis of Integral Membrane Proteins
- Proof of Principle

- Influenza A virus M2 H⁺ Channel
  - tetramer, 97aa polypeptide

Kochendoerfer et al.
Biochemistry 38, 11905-13 (1999)
A ‘One-Pot’ Total Chemical Synthesis of Crambin

Bang & Kent,
Characterization of Synthetic Crambin

MS of folded Crambin:

Calculated: 4702.41 Da
Observed: 4702.2 ± 0.3 Da
Crambin Synthesis - yields

0.4 millimole scale SPPS

1. 

\[
\begin{align*}
\text{SH} & \quad \text{SH} \\
& \quad 1 \quad 15 \\
& \quad \text{CO-SR} \\
& \quad 280 \text{mg (56%)}
\end{align*}
\]

2. 

\[
\begin{align*}
\text{SH} & \quad \text{SH} \\
& \quad 1 \quad 16 \\
& \quad \text{Thz} \\
& \quad 356 \text{mg (48%)}
\end{align*}
\]

\[
\begin{align*}
\text{SH} & \quad \text{SH} \\
& \quad 31 \quad \text{CO-SR} \\
& \quad \text{Cys} \\
& \quad 340 \text{mg (49%)}
\end{align*}
\]

3. 

\[
\begin{align*}
\text{SH} & \quad \text{SH} \\
& \quad 16 \quad \text{Cys} \\
& \quad \text{ligate} \\
& \quad \text{deprotect (66%)}
\end{align*}
\]

4. 

\[
\begin{align*}
\text{SH} & \quad \text{SH} \\
& \quad 46 \quad \text{COO}^- \\
& \quad \text{fold/disulfides (88%)}
\end{align*}
\]

5. 

\[
\begin{align*}
\text{SH} & \quad \text{SH} \\
& \quad 40 \quad 26 \quad \text{Crambin (20% based on starting resin)}
\end{align*}
\]

6. 

\[
\begin{align*}
\text{SH} & \quad \text{SH} \\
& \quad 32 \quad \text{Crambin (42% based on peptides)}
\end{align*}
\]

\[
\begin{align*}
\text{SH} & \quad \text{SH} \\
& \quad 46 \quad \text{COO}^- \\
& \quad 375 \text{mg}
\end{align*}
\]
Total Synthesis of *human* G-Protein Coupled Receptors

- Large proteins (>350 residues)
- Glycoproteins
- Xray structures
- Single molecule fluorescent studies
  - functional states/signal transduction
- Drug targets - mirror image drug discovery
Emil Fischer (1852 - 1919)

- 'Lock & Key' theory of enzyme action (1894)
- 'Peptide theory' of protein structure (1902) [F. Hofmeister]

"My entire yearning is directed toward the first synthetic enzyme. If its preparation falls into my lap with the synthesis of a natural protein material, I will consider my mission fulfilled."

E Fischer
(letter to Adolf Baeyer, 5th Dec 1905)
Modern Chemical Protein Synthesis

peptide synthesis \[\text{plus}\] chemical ligation

\[
\begin{align*}
\text{large polypeptides} & \quad \text{unprotected peptides} \\
& \quad \text{chemoselective rxn}
\end{align*}
\]

Key technologies:
- solid phase peptide synthesis
- native chemical ligation (NCL)
- electrospray MS
- protein NMR
- Xray crystallography

\[
\begin{align*}
\text{protein molecule} & \quad \text{fold/disulfides} \\
& \quad \text{defined 3\textdegree structure} \\
& \quad \text{biological function}
\end{align*}
\]
Native Chemical Ligation

N-Terminal Segment → water, pH 7 → C-Terminal

unprotected peptide segments

amide-forming chemoselective rxn

Unique Properties of Thioester Functionality

Thioester, -COSR

- stable to hydrolysis (H₂O/OH⁻)
- labile to RS⁻
- labile to R-NH₂

10,000-fold

rxn conditions: aqueous 6M GnHCl, pH7

freely reversible

irreversible

Dawson, Muir, Clark-Lewis, & Kent
Science, 266, 776 (1994)
Pancreatic RNase A [124 aa; 8 Cys - 4 disulfides]

- Crystallized with JG365
- Space group is $P2_12_12_1$
- Data set 1.04Å resolution

Sequential ligation

$13\,639 \pm 2\text{Da}$
Convergent Synthesis

- minimum # steps, time, effort
- minimum amounts of starting materials
- efficient & versatile analogue synthesis

Systematic Synthesis Design. 6. Linear Yield Analysis and Convergency
(Partially Convergent)

James B. Hendrickson

Journal of the American Chemical Society / 99. 5439 (1977)
Convergent Chemical Ligation
Convergent Chemical Ligation

Challenge: independent control of reactivity of each end
Kinetically controlled ligation

NO thiol catalyst

Conversion of Thz- to Cys- in presence of thioester

Bang & Pentelute
Angew. Chem. (2006)
Key Cys-peptide-thioester intermediates

Extend at either end - at will!
Inventing New Chemistries to Reveal How Proteins Work

**Chemical Protein Synthesis**

- Atom-by-atom control over molecular structure
- Unlimited sites/number/kinds of 'non-coded' analogues

**Enables**

- Exquisitely precise, systematic dissection of the molecular basis of protein function

**Using**

- Physical techniques (FTIR, nmr, Xray crystallography, DSC, protein MS, etc.)
Chemical Protein Analogues

1. Non-coded amino acids

2. Mirror image proteins
   *Science, 256*, 1445 (1992)

3. Protein diastereomers
   *Nature Chemical Biology, 2*, 139 (2006)

4. Fixed elements of 2° structure
   *Protein Science*, 2, 1085 (1993)

5. Backbone engineering
   *PNAS(U.S.A), 90*, 11638 (1993)

6. Topological analogues

7. Site-specific labeling
   *Nature Structural Biology, 3*, 946 (1996)

8. Glycoprotein mimetics
Current Kent Lab Research

Applying chemistry to:

- Chemistry of enzyme catalysis
- Integral membrane proteins
- Topological analogues
- Racemic protein crystallography
- Mirror image Rx discovery
- Chemical manufacture - human insulin
- Nanoribbons (Shohei Koide)
- Parallel chemical protein synthesis
Role of 'flaps' in enzyme catalysis?
Convergent Synthesis HIV-1 PR 'Covalent Dimer'

PQITLWKRPL VPITIRIGGQLK EALLDTGADD TVIEEMNLPG KWKPKMIGGI
GGFIKVQRQYD QQIPEIXGHIK AIGTVLVGPT PVNIIGRNLL TQIGXTLNFC
GGGPGQITLW KRPLVTIRIG GQLKEALLDT GADDTVIEEM NLPGKWKPMIGGI
IGGIGGFIKV RQYDQIPVEI XGHKAIITVL VGPTPVNIIG RNLLTQIGXTLNF

203 amino acids

* MPAA - mercaptophenylacetic acid thioester
** KCL - kinetically-controlled ligation
*** Alkylation with 2-bromoacetamide

Torbeeve
Characterization of HIV-1 PR 'covalent dimer' construct

Observed mass 21869.8 ± 0.4 Da
[calc: 21869.76 (av. isotopes)]
**KINETIC DATA:** \( \text{Gly}^{51}, \text{Gly}^{151} \rightarrow \text{Xaa}^{51}, \text{Yaa}^{151} \) analogues

<table>
<thead>
<tr>
<th>Type</th>
<th>Gly(^{51})</th>
<th>Gly(^{151})</th>
<th>(k_{\text{cat}}) (s(^{-1}))</th>
<th>(K_m) (µM)</th>
<th>(k_{\text{cat}}/K_m)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild type</strong></td>
<td></td>
<td></td>
<td>23.4</td>
<td>25</td>
<td><strong>0.93</strong></td>
</tr>
<tr>
<td><strong>Symmetric analogues</strong></td>
<td>(S)-Ala(^{51})</td>
<td>(S)-Ala(^{151})</td>
<td>3.7</td>
<td>50</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(R)-Ala(^{51})</td>
<td>(R)-Ala(^{151})</td>
<td>4.9</td>
<td>434</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Aib(^{51})</td>
<td>Aib(^{151})</td>
<td>(very low activity)</td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Asymmetric analogues</strong></td>
<td>(R)-Ala(^{51})</td>
<td>(S)-Ala(^{151})</td>
<td>17.6</td>
<td>26</td>
<td><strong>0.67</strong></td>
</tr>
<tr>
<td></td>
<td>Gly(^{51})</td>
<td>(S)-Ala(^{151})</td>
<td>22.2</td>
<td>47</td>
<td><strong>0.47</strong></td>
</tr>
<tr>
<td></td>
<td>Gly(^{51})</td>
<td>(R)-Ala(^{151})</td>
<td>6.2</td>
<td>44</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Gly(^{51})</td>
<td>Aib(^{151})</td>
<td>4.0</td>
<td>105</td>
<td>0.04</td>
</tr>
</tbody>
</table>
X-ray Structures of Chemical Analogues of HIV-1 Protease
All isomorphous in $P_{2_1}2_12_1$; ligand MVT-101 (Ac-Thr-Ile-Nle-(CH$_2$NH)Nle-Gln-Argamide)
X-ray Structures of Chemical Analogues of HIV-1 Protease
All isomorphous in $P2_12_12_1$; ligand MVT-101 (Ac-Thr-Ile-Nle-(CH$_2$NH)Nle-Gln-Argamide)

1.3 Å resolution

1.55 Å resolution
Based on:
D.B. Northrop
Acc. Chem. Res. 34, 790-797 (2001)
Direct Observation of ‘Low-Barrier H-Bond’ Proton

1H NMR (chemical shift (ppm))
Ongoing studies:

- Confirm 'low barrier' H-bond - [4-$^{13}$C]Asp$^{25}$ PR (J-coupling)
- Site-specific {EPR+fluorophore labels} - dynamic distances
- Site-specific NMR labels - flap dynamics in analogues
- Single molecule {force/fluorescence} measurements - state of the flaps in act of catalysis
TIRF Single Molecule Studies of HIV-PR

Upon addition of substrate

Torbeev 2007
Challenges in Chemical Synthesis of Integral Membrane Proteins

- large polypeptides
  - ‘one pot’ synthesis
  - convergent synthesis
- TM peptide-thioesters
  - synthesis
  - solubility & handling
- solubility & handling of intermediates
  - polymer-supported chemical ligation
- folding
Major Issue - TM Peptide Solubility

Diacylglycerol kinase

Major Issue - TM Peptide Solubility

Diacylglycerol kinase

b) 26-45

\[ \text{Thz-NEAAFRQEGVAVLLAVVIA-COSCH}_2\text{CH}_2\text{CO-X} \]

- \( X = \text{Gly} \)
- \( X = \text{Arg}_6 \)

[Graphical representation of peptide solubility]
Major Issue - TM Peptide Solubility

Diacylglycerol kinase

d) 61-83

\[ Thz-VMLVMIVEILNSAI\text{EAVVDRIG-COSCH}_2\text{CH}_2\text{CO-X} \]

\[ X = \text{Gly} \]

\[ X = \text{Arg}_6 \]

e) 84-112

\[ Thz-EYHELSGRAKDMGSAAVLIAIIVAVITW-COSCH}_2\text{CH}_2\text{CO-X} \]

\[ X = \text{Gly} \]

\[ X = \text{Arg}_6 \]

Total synthesis of a human GPCR

Extracellular

Intracellular
Human delta-Opioid Receptor

Peptide 6

\[ \text{CWAPIHIF}^{280} \text{VIVWTLDID}^{290} \text{RRDPLWAAL}^{300} \text{HL-COSR-KKKKK} \]

a) 23 residues

b) 37 residues

Erik Johnson.
• only 65/230 space-groups are **chiral**

• P1(bar) most favored space-group, **achiral**

• **Prediction**: racemic protein mixtures will crystallize more readily!

Glycine-Rich Antifreeze Proteins from Snow Fleas

Laurie A. Graham¹ and Peter L. Davies¹,2*

SCIENCE VOL 310 21 OCTOBER 2005

Snow flea antifreeze protein ('sfAFP'):
• 81 residues
• 2 disulfides
• thermally unstable
• hard-to-express by recDNA
Total Chemical Synthesis of sfAFP

Inhibition of ice crystal growth

1) buffer alone
2) ~25 mg/ml AFP I,
3) ~25 mg/ml sfAFP (SH)$_4$
4) ~25 mg/ml sfAFP.

Pentelute, 2007
Crystallization of sfAFP

- difficult, badly twinned
- no crystals from Se-sfAFP

sfAFP
38 mg/mL

sfAFP
19 mg/mL
Racemic Crystallization of sfAFP

Crystals were obtained in >50% of conditions!

A. (D + L)-sfAFP
   38 mg/mL

B. sfAFP
   38 mg/mL

C. sfAFP
   19 mg/mL
Acknowledgements

• DOE 'Genomes to Life'
• NIH Roadmap
• NSF/MRSEC
• Physics, Chemistry, Biology undergrad research program

Support from . . .

• DOE 'Genomes to Life'
• NIH Roadmap
• NSF/MRSEC
• Physics, Chemistry, Biology undergrad research program

Duhee Bang*
Alex Donovan
Thomas Durek*
Zachary Gates*
Erik Johnson*
Suhai Liu
Kalyan Mandal
Laurie Parker
Brad Pentelute
Youhei Sohma
Vivian Tien
Vladimir Torbeev

* graduated/left

Collaborators
Tony Kossiakoff & Valya Terechko (Chicago)
Taekjip Ha (UI-UC)