Chapter 9: Polyamide/NCP Ligation and Profluorescent Azido-Carbostyrils

Abstract

The nucleosome core particle (NCP) represents a physiologically relevant fundamental repeating unit of chromatin. The ability to modulate the structure of chromatin using sequence specific DNA binding small molecules could have potential applications in gene regulation, biotechnology, physical biology, and macromolecular structure studies. The possibility of self-assembling small molecules inside living cells using biological architectures represents a novel approach to molecular recognition. Small molecule probes may be used to report on dynamic information, regulate structural changes, and provide stabilization for structure determination of inherently dynamic macromolecules. Sequence-specific polyamides are a modular platform and could be used as DNA-binding building blocks to template the self-assembly of larger polyamides that target higher order structures such as the NCP, the tetra-NCP, the chromatin fiber, or other diverse nanometer scale structures inside the cell. This study provides a proof-of-principle experiment demonstrating that polyamides can be self-assembled using a high-order nucleoprotein structure (NCP) as a template.

9.1 Introduction

9.1.1 Templated Dimerization of Polyamides

The ability to use double stranded DNA as a template for the ligation of two polyamides with bioorthogonal functionality has recently been demonstrated.[1,2] The 1,3-dipolar cycloaddition between an azide and alkyne was employed, resulting in a triazole linked tandem hairpin polyamide capable of targeting >10 base pairs of DNA with rate increases of 20,000-fold.[1,2] The ability to template the dimerization of two cell permeable molecules for the targeted downregulation of a gene represents a powerful strategy for the control of transcriptional regulated processes in biological systems. The Dervan laboratory has also shown that polyamides are capable of binding the nucleosome core particle (NCP) and that a polyamide clamp constructed from two 8-ring polyamides connected at the turns by a linker is capable of locking the DNA onto the histone octamer (Figure 9.1).[3,4] This NCP clamp was able to lock a complete turn of DNA onto the octamer significantly improving *in vitro* stability. Unfortunately, NCP clamps are very high molecular weight branched oligomers (MW > 2500) and often have poor cell permeability profiles.[5,6] The ability to template the dimerization of two hairpin polyamides across the supergroove between the two gyres of DNA on the NCP would allow for the introduction of two lower molecular weight polyamides potentially obviating cell permeability and size issues.

Incorporation of reporting strategies into the ligation event such as a profluorescent linker, where fluorescence would be activated by the dimerization process, would allow for the direct monitoring of reaction kinetics and cellular localization. The clamp dimerization strategy is illustrated in Figure 9.2 and would provide an important tool for the study of nucleosome core
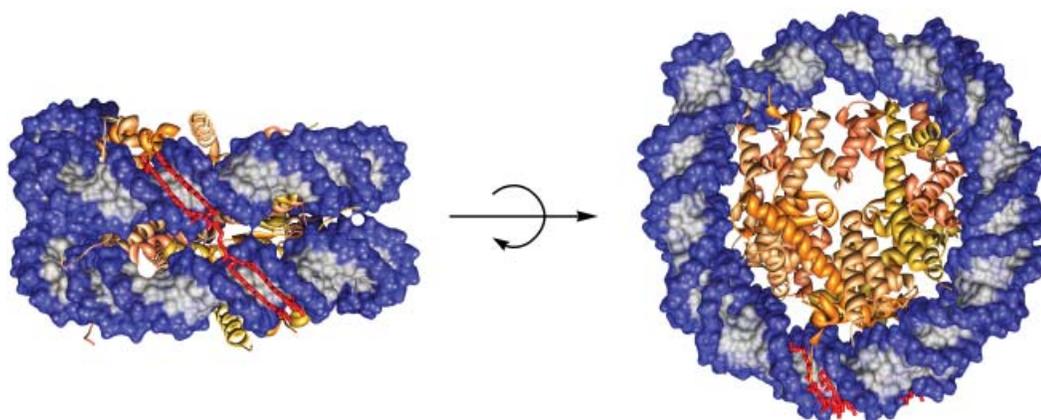
Figure 9.1 High resolution crystal structure of an NCP bound polyamide clamp. Clamp is shown in red. DNA backbone shown in blue with minor and major grooves in gray. Side view (left) showing histone octamer core (orange, yellow, tan, and pink). Top view (right) with clamp bound in a minor supergroove on the proximal face. (PDB 1S32)

particle recognition and gene regulation in addition to serving as a macromolecular structural probe. The potential for gene regulation at the NCP level represents a unique physiologically relevant recognition platform distinct from that of linear DNA. In addition, NCP's tend to be sensitive to radiation damage and this along with dynamic DNA dissociation significantly reduce crystal diffraction quality.[4] Polyamide clamps provide a unique tool for structural biology due to their ability to increase the size, order, and resolution of NCP crystals, as evidenced by the 2 Å resolution structure in Figure 9.1, providing the potential for crystallization of larger physiologically relevant structures.[4] One could dream of the possibility of crystallizing structures such as multiply linked histone octamers or NCP's with bound transcription factors and clamps of this sort could be just the
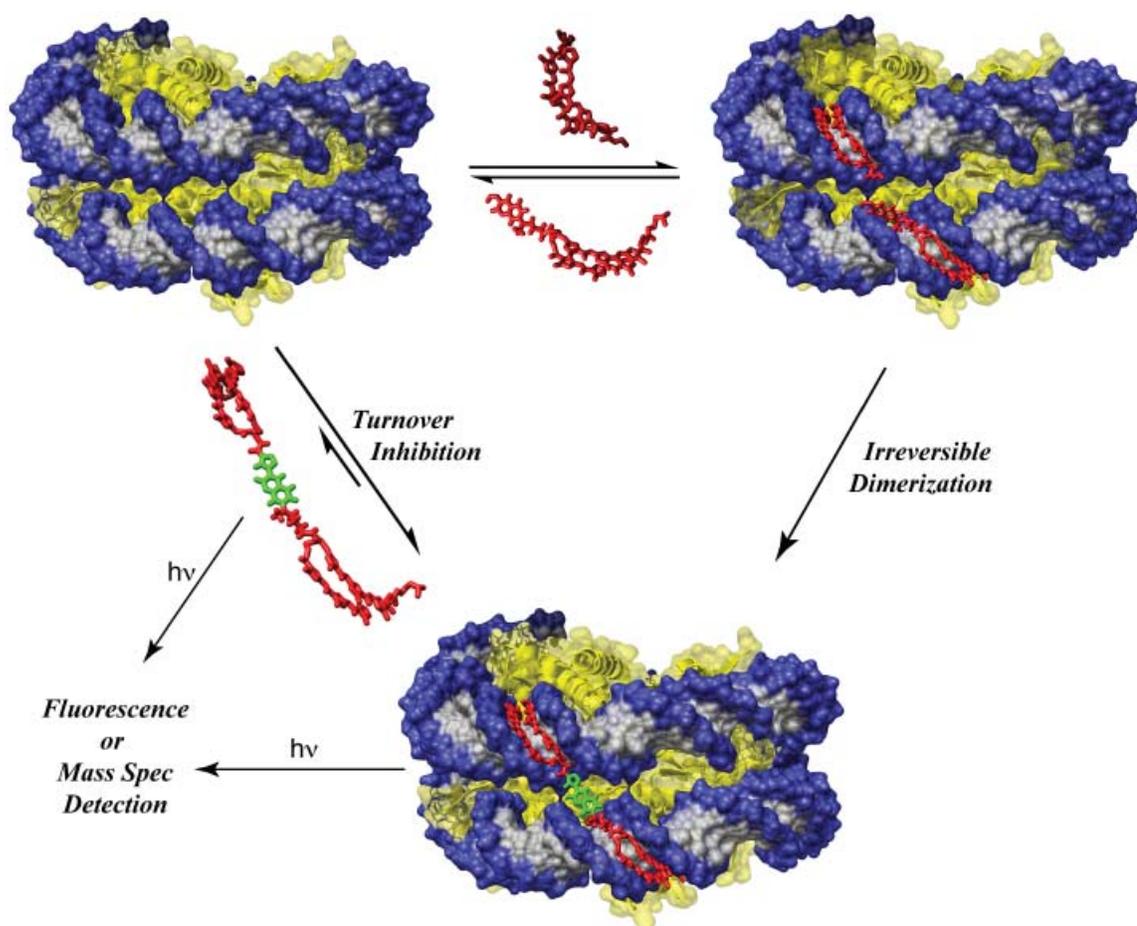
Figure 9.2 Illustration of the overall clamp dimerization strategy. Nucleosomal DNA shown without histone octamer present for clarity. Initial polyamide binding is a reversible process along with unwinding and winding of the DNA from the histone octamer. Dimerization of the two polyamides via a bioorthogonal reaction templated by the NCP is irreversible resulting in a large, sequence specific, high affinity clamp. The clamp can serve to stabilize the NCP for structural studies or potentially downregulate gene expression. Direct kinetic information and cellular localization data can be obtained via dimerization induced profluorescent linker activation.

tool needed to gain structural insight into these complex biological architectures.
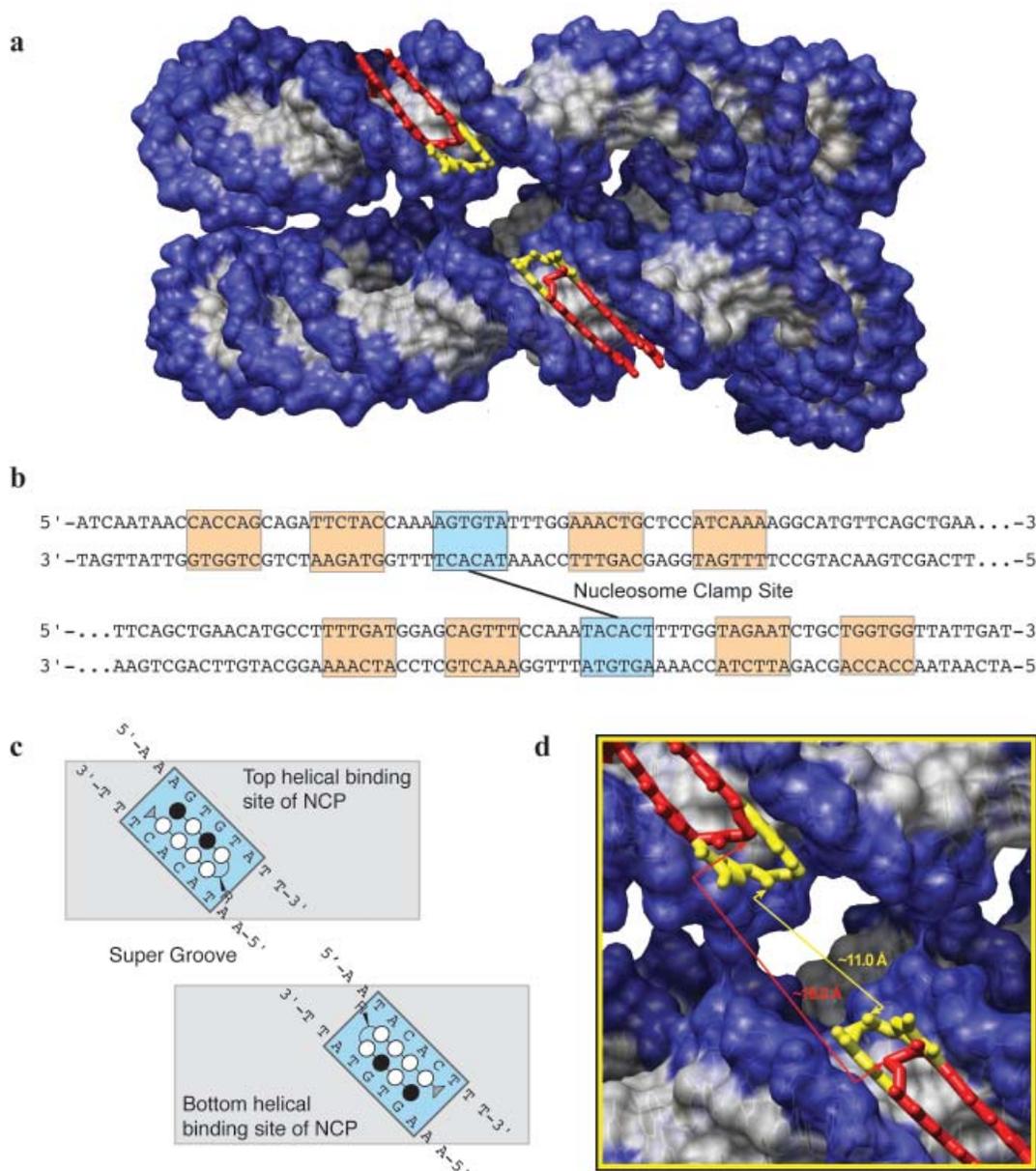
Figure 9.3 a) Sequence of the 146 base pair fragment of α-satellite DNA used for crystallographic studies with the nucleosome clamp. In those structures, the clamp was found to bind in the homodimeric "supergroove" highlighted in yellow. Each of the other four supergrooves on the NCP are highlighted in purple, green, blue and red. a) Highlight of the sequences to which the nucleosome clamp was bound (left). At right is shown the two base pair mutation (in red) introduced so that the supergroove becomes heterodimeric. Located above each highlighted site is the polyamide designed to target that site. b) Analysis of the linker dependence of NCP templated ligations. Crystal structure view of the supergroove. The predicted distance between the amines of the turn are shown in green for binding at the original site, and in red if the sites are moved one base pair back. c) Modeling of the linkers and the calculated turn-to-turn length.

9.2 Results and Discussion

In order to ascertain whether the NCP could be used to template ligation reactions (Figure 9.3), a series of three azide containing polyamides and two alkyne containing polyamides were synthesized. Several linker lengths were used in order to examine the distance dependence of the reaction (Figure 9.4). Examination of the previous crystal structure data and computational modeling
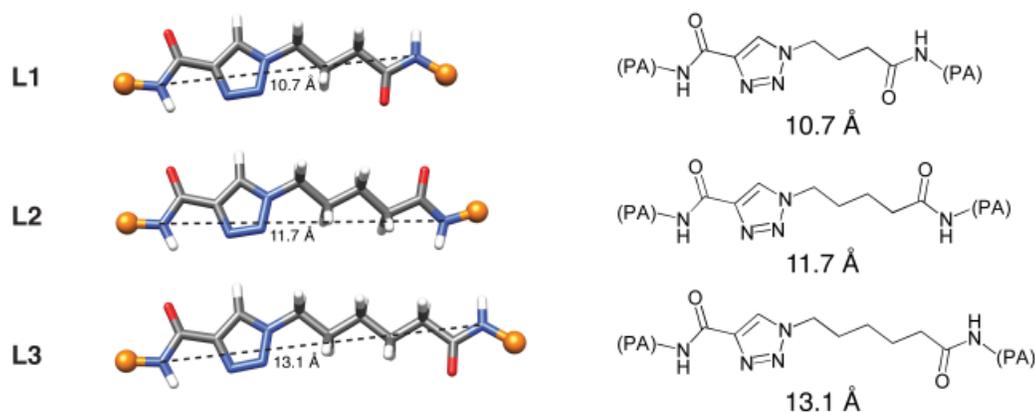
Figure 9.4 Analysis of the linker dependence of NCP templated ligations. Models and structure of the linkers with the calculated turn-to-turn length listed below each.

showed that only the azide containing the longest linker was expected to be capable of reacting. In addition, the previous study of DNA ligation had demonstrated that alkyne had reacted over 20,000 times faster than the alkyne containing an extra carbon unit and was expected to perform similarly in our experiments. The alkynes were purchased commercially and then coupled directly to the appropriate polyamides as described previously. The azides were synthesized using the scheme shown in Figure 9.5, 9.6, and 9.7.

The NCP was reconstituted as outlined in the supplemental information. A control sample containing only the 146 bp DNA and a second control containing
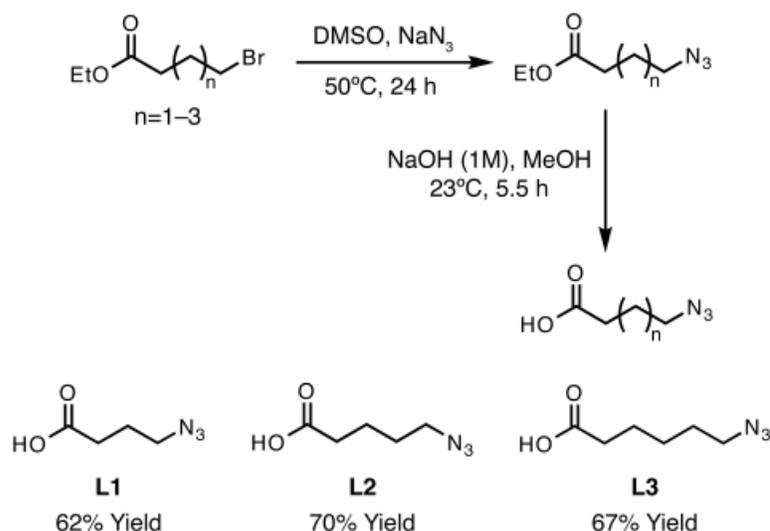
Figure 9.5 Synthesis of alkyl azido linkers.[6]

only the histone octamer were prepared as well. Next, 200 pmol each of **4** and **5** was added to 40 pmol of the reconstituted NCP and the reaction was incubated for 5 h at 37°C. MALDI-TOF MS was used to analyze the reaction, and after 5 h peaks were observed corresponding to the expected mass of the ligation product ([M+Na]$^+$ at 2687.77 and [M+H]$^+$ at 2664.88, which is in agreement with the calculated mass of 2664.23) (Figure 9.8). Control samples were run under the same conditions containing either just the DNA, just the histone octamer, or only buffer. In all three cases the target mass was not observed upon addition of the polyamide azide and alkyne, indicating that the ligation reaction was dependent upon the presence of the fully reconstituted NCP (Figure 9.9). The lack of product observed in the DNA only control reaction is promising, indicating there is reaction specificity dependent on DNA geometrical constraints imposed by assembly of the NCP. This is a DNA binding-site proximity-based reaction, dependent on the proper assembly of a higher order macromolecular structure with complete specificity over linear DNA. This study provides evidence that our reaction is taking place in a templated format and the sequence specificity of the polyamide reagents is ensuring that the reaction only takes place when the two targeted binding site sequences are properly aligned in 3-dimensional space.

The same ligation reaction was performed using all three azides and both alkynes. Of the six potential ligation reactions, product was only observed in the case of compounds **4** and **5**. This result agrees with both computational work indicating the linkers for compounds **2** and **3** were too short to react, as well as prior work demonstrating that the alkyne used in **6** was significantly less reactive.

These results demonstrate the ability to perform templated ligation reactions on the NCP as confirmed by MALDI-MS analysis, however the small scale of the reaction prevented
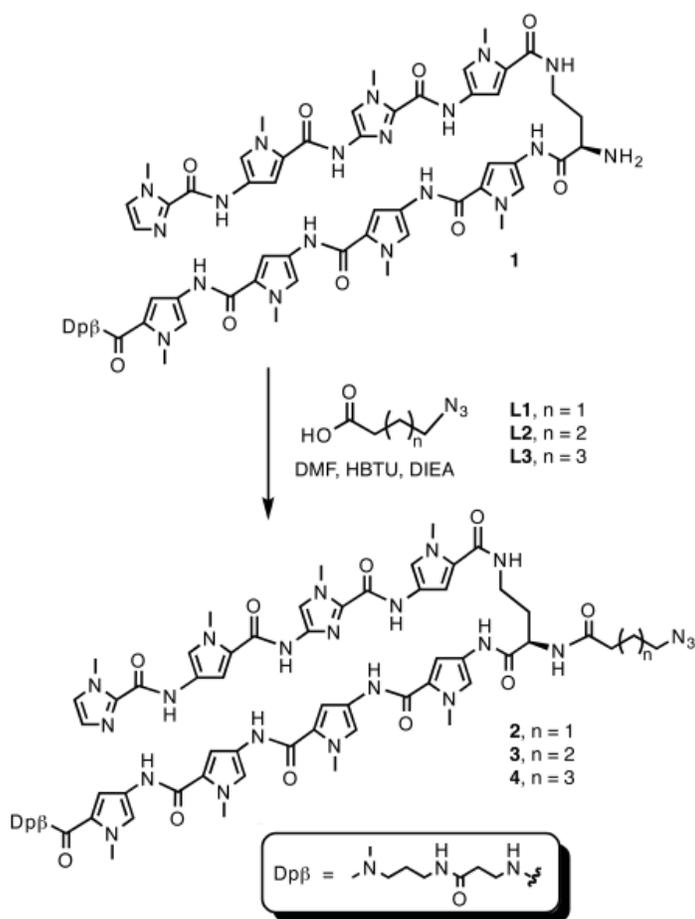
Figure 9.6 Synthesis of azido-polyamides.

the quantitation of ligation yield. As previously discussed, a fluorescent reporter strategy utilizing pro-fluorescent probes could provide an alternative to MALDI-MS analysis, with far less background noise.

Coumarins have enjoyed widespread use as platforms for the discovery of fluorescent molecules (e.g. laser dyes). Reports of profluorescent coumarins, which upon chemical modification or reaction give rise to a fluorescent coumarin based molecules have been reported.[7] Azido coumarins have been shown to react with alkynes via 1,3-dipolar cycloadditions giving rise to highly fluorescent products due to a change in donor acceptor properties of appended functionality (Figure 9.10).[7] In addition, maleimide-functionalized coumarins are similarly non-fluorescent until conjugate addition with a thiol yields a fluorescent product.[8] These reactions could potentially be used in a bioorthogonal profluorescent ligation strategy for templating the dimerization of polyamides on the nucleosome core particle as illustrated in Figure 9.2. The ability of the linker to act as a fluorescent switch offers the unique possibility of monitoring reaction progress and cellular localization using highly sensitive fluorescent techniques. To establish linker requirements for the coumarins, molecular modeling was done using Spartan ES software package[9] and energy-minimized using an AM1 model, followed by ab initio calculations by means of the Hartree-Fock method using the 6-31G* Pople basis set.[10] As shown in Figure 9.10, the fully compacted, shortest triazole-coumarin spans 13 Å, which is 2 Å longer than the 11.0 Å distance between the two 8-ring polyamide turn amines in the crystal structure (Figure 9.3). This posed a potential problem due to the rigidity of the of the coumarin system. When 6-ring hairpin polyamides were modeled in for the 8- ring polyamides on the

Figure 9.7 Synthesis of alkynyl-polyamides.

crystal structure, the distance between turn amines increased to 17.8 Å. Replacement of 8-ring polyamides for 6-ring polyamides was a potential solution to the problem, but not ideal since all structural data to date has been generated for 8-ring polyamides bound to the NCP. After surveying the literature for potential profluorescent replacements the aza-analogues of coumarins (carbostyrils) were discovered. Carbostyrils[4,11] exhibit photophysical properties similar to coumarins and can

Azido-Polyamide	Linker Length	(PA)⌒≡ **5**	(PA)⌒≡ **6**
2	n=1	N.R.	N.R.
3	n=2	N.R.	N.R.
4	n=3	Product (7) calcd. mass: 2664.23 found: [M+H]⁺ 2664.88 found: [M+Na]⁺ 2687.77	N.R.

PA = Polyamide
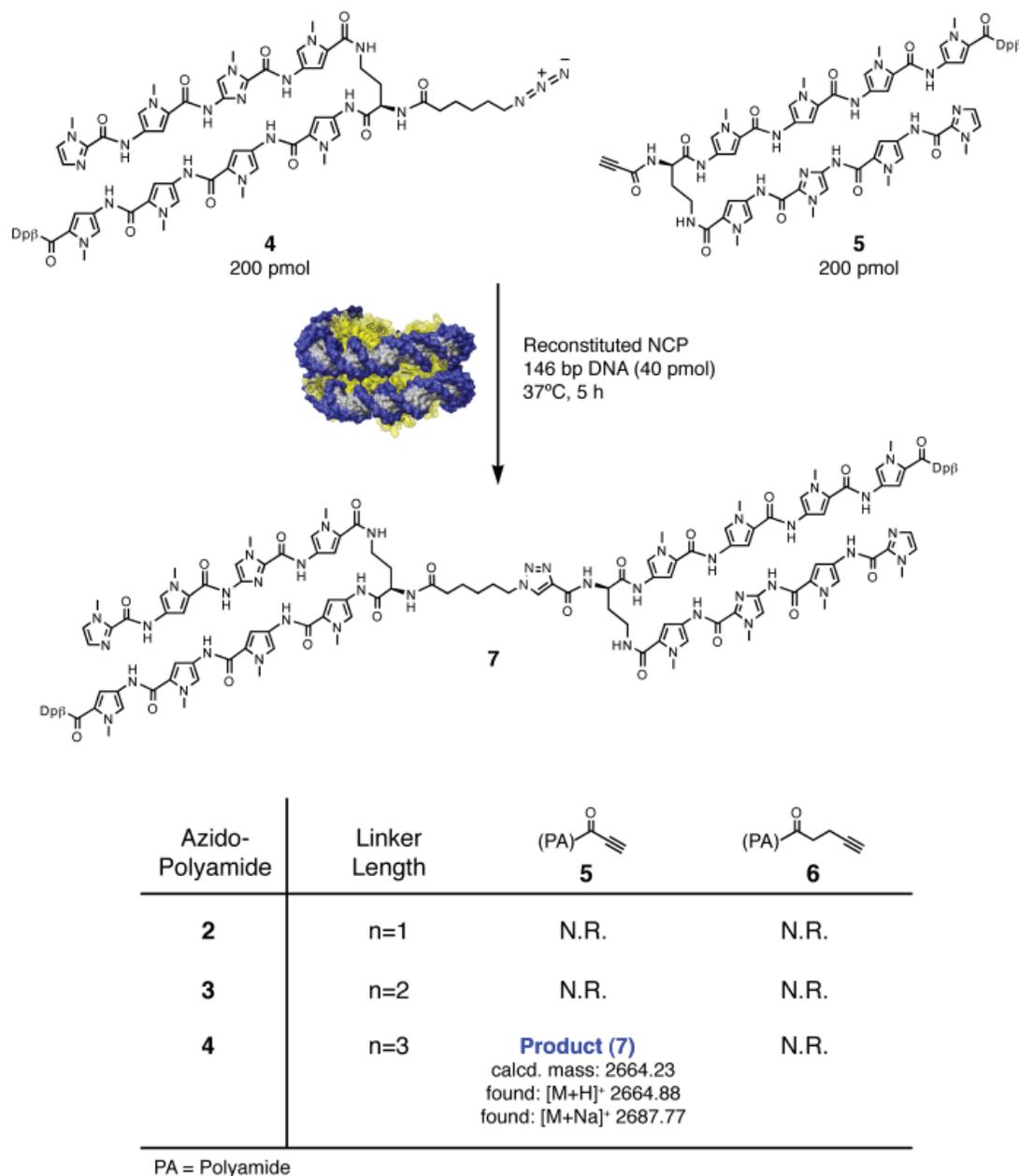
Figure 9.8 DNA templated ligation on the NCP. The reaction of **4** and **5** with the NCP leads to dimer **7** which was observed by MALDI mass spectrometry. NCP templated ligation between different length azides and alkynes. The reaction product was only detected with polyamides **4** and **5**. N.R. = no reaction.
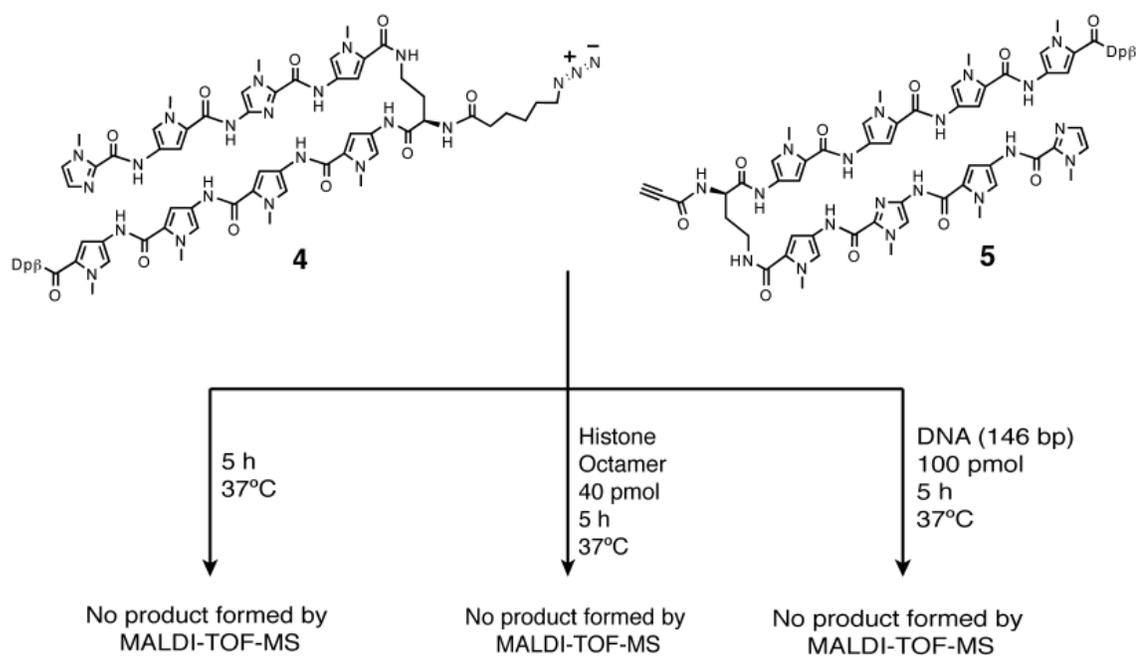
Figure 9.9 Control reactions for the NCP templated ligation. The ligation reaction does not occur in the presence of just buffer, just the histone octamer, or just 146 bp DNA.

be electronically tuned by substitution with electron withdrawing and donating functionality.[8,12] Figure 9.11 shows the structure and linker distance comparison for carbostyrils versus coumarins. It appears that the carbostyril based linkers will be able to accommodate both 6- and 8-ring systems. In addition, synthetic ease, compact size, and a nitrogen handle for derivatization makes the carbostyrils an ideal candidate for use as a coumarin alternative. Preliminary efforts towards the synthesis of carbostyril (**17**) are presented in Figure 9.12.

Synthesis of carbostyril (**17**) (Figure 9.12) started with condensation of commercially available *N,N*-dimethyl-m-phenylenediamine (**8**) with dimethyl malonate at 200 °C to give 2,4-dihydroxy-7-(dimethylamino)quinoline (**14**) in 50% isolated yield after a single recrystallization. Chlorination of **14** under refluxing POCl$_3$ afforded 2,4-dichloro-7-(dimethylamino)quinoline (**15**) in 80% yield following recrystallization. Quinoline **15** was hydrolyzed under refluxing conditions in 6M HCl to afford 4-chloro-7-(dimethylamino)carbostyril (**16**) in 90% yield. Selective hydrolysis of the chlorine at position 2 is often explained by the observation that the 2 position is more reactive toward nucleophilic substitution than the 4 position. After comparing the proton NMR of compound **16** to that reported in the literature, an apparent discrepancy was realized. Since there was a possibility of displacement of chlorine at two different positions and NMR would not unambiguously resolve the regioselectivity issue, the compound was recrystallized from hot DMF
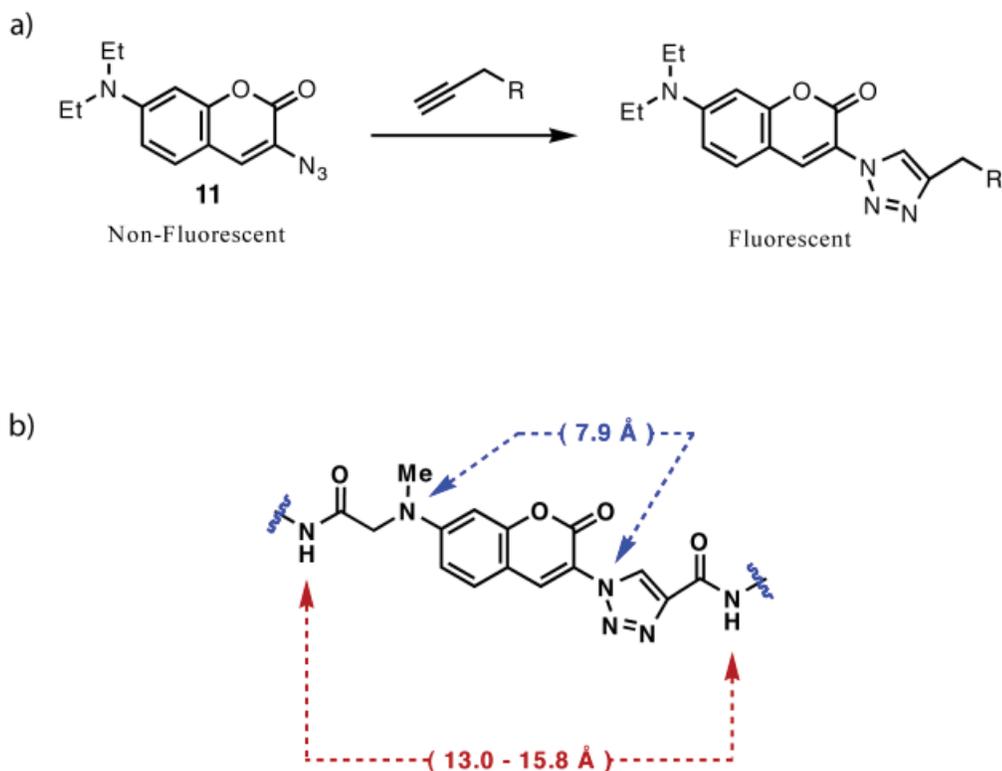
Figure 9.10 a) Reaction of a profluorescent azido coumarin giving rise to a fluorescent triazole product. b) Linker distances for incorporation of the profluorescent coumarin.

and the crystal structure solved. The crystal structure and unit cell are shown in Figure 9.12. After unambiguously proving the identity of compound **16** it was subjected to azidification. Compound **16** was treated with excess sodium azide in d_6-DMSO and the reaction monitored by proton NMR. After heating for exactly 10.5 h at 120 °C complete conversion to azide **17** was achieved by NMR. Compound **17** can be N-alkylated to provide a functional handle for conjugation to small molecules and macromolecules. Initial investigations into the photophysical properties of compound 17 are shown in Figure XX and demonstrate the ability of the probe to fluoresce upon cycloaddition with an alkyne substrate.

9.3 Conclusion

These studies have demonstrated the feasibility of the NCP templated ligation approach for the self assembly of polyamide dimers. Additionally, the development of a pro-fluorescent azido carbostyril provides a new tool for monitoring ligation reactions using fluorescence. This strategy offers an exciting opportunity for modifying gene expression in cells by the targeted self-assembly of polyamides on NCP's.
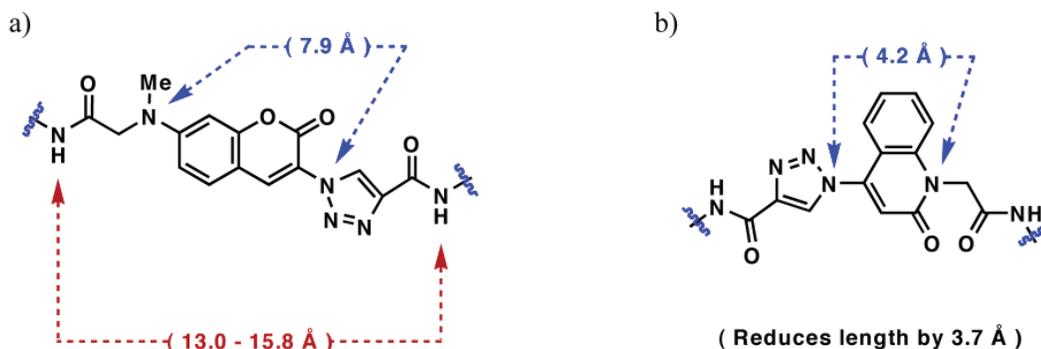
Figure 9.11 a) Coumarin linker length presented for comparison against carbostyril. b) Carbostyril linker distance shows a decrease in length of ~3.7 angstroms.

9.4 Experimental

9.4.1 Materials and General Methods

Dicyclohexylcarbodiimide (DCC), 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), *N*-Hydroxybenzotriazole (HOBt), Fmoc-Dab(Boc)-OH and Boc-β-Ala-Pam resin were purchased from Peptides International. Benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) was purchased from Novabiochem. Flouro-*N,N,N',N'*-tetramethylformamidinium hexafluorophosphate (TFFH) was purchased from Advanced ChemTech. O-(7-Azabenzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HATU), 4-(dimethylamino)-pyridine (DMAP), *N*-hydroxysuccinimide (NHS), *N,N*-dimethylformamide (DMF),*N*-methylpyrrolidinone(NMP),*N,N*-dimethylpropylamine(Dp),*N,N*-diisopropylethylamine (DIEA), ethylene diamine, piperidine, and other miscellaneous chemicals were purchased from Aldrich and used without further purification. All other solvents were purchased from EM Sciences and were reagent grade. Trifluoroacetic acid (TFA) was purchased from Halocarbon.

^{1}H NMR spectra were recorded using a 300MHz General Electric-QE NMR spectrometer. CDCl$_3$ was obtained from Cambridge Isotope Laboratories. UV spectra were recorded in water using a Beckman Coulter DU 7400 Spectrophotometer. Matrix-assisted LASER desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) was performed using an Applied Biosystems Voyager DE Pro Spectrometer. Electrospray ionization (ESI) mass spectrometry was performed by the Protein and Peptide Microanalytical Facility at the California Institute of Technology. Analytical High-Pressure Liquid Chromatography (HPLC) was performed with a Beckman Gold system using a Varian Microsorb-MV 100 C18 column (5μm particle size, 250 x 4.6mm). Preparative HPLC was performed using a Beckman Gold system with either a Waters Bondapak C18 column (15-

a)

b)

16
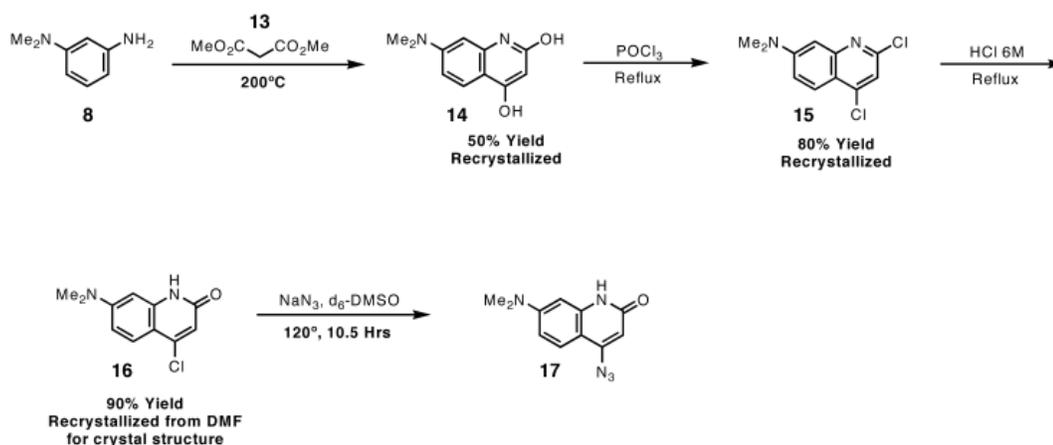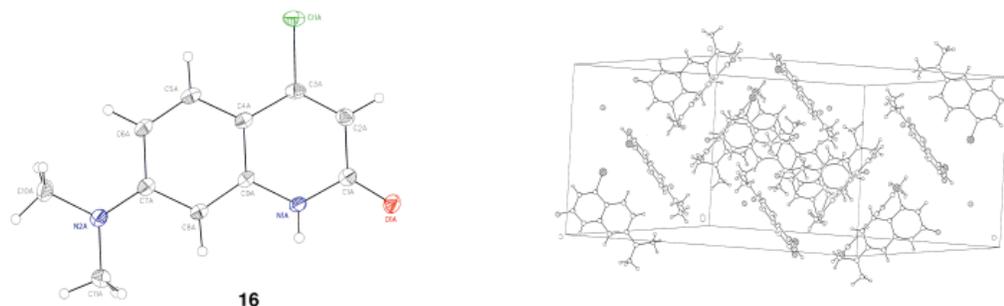
Figure 9.12 a) Progress toward the synthesis of profluorescent azide **18**. b) X-ray crystal structure and unit cell for structural proof of compound **16**.

20 μm particle size, 25 x 100mm) or a Phenomenex Gemini C18 column (5 μm particle size, 250 x 21.2 mm). For both HPLC systems Solvent A was 0.1% (v/v) aqueous TFA and solvent B was acetonitrile. Analytical HPLC was done using a gradient of 1.85%/min of Solvent B starting from 0% over 35 min with a flowrate of 1.5mL/min. Preparative HPLC was typically done using a gradient of 1%/min of Solvent B for 20 min followed by a gradient of 0.3%/min for an additional 100min at a flowrate of 8mL/min. Radioactive gels were imaged using a Molecular Dynamics 400S PhosphorImager.

Restriction endonucleases, deoxyribonucleotide triphosphates, DNAse I, Polynucleotide kinase (PNK), and glycogen were purchased from Roche. [α-^{32}P]-Thymidine-5'-triphosphate (≥ 3000 Ci/mmol) and [α-^{32}P]-Deoxyadenosine-5'-triphosphate (≥ 6000 Ci/mmol) were purchased from Perkin Elmer. [γ-^{32}P]-Adenosine-5'-triphosphate (≥ 7000 Ci/mmol) was purchased from MP Biomedicals. Water was purified from a Millipore Mill-Q purification system for general use.
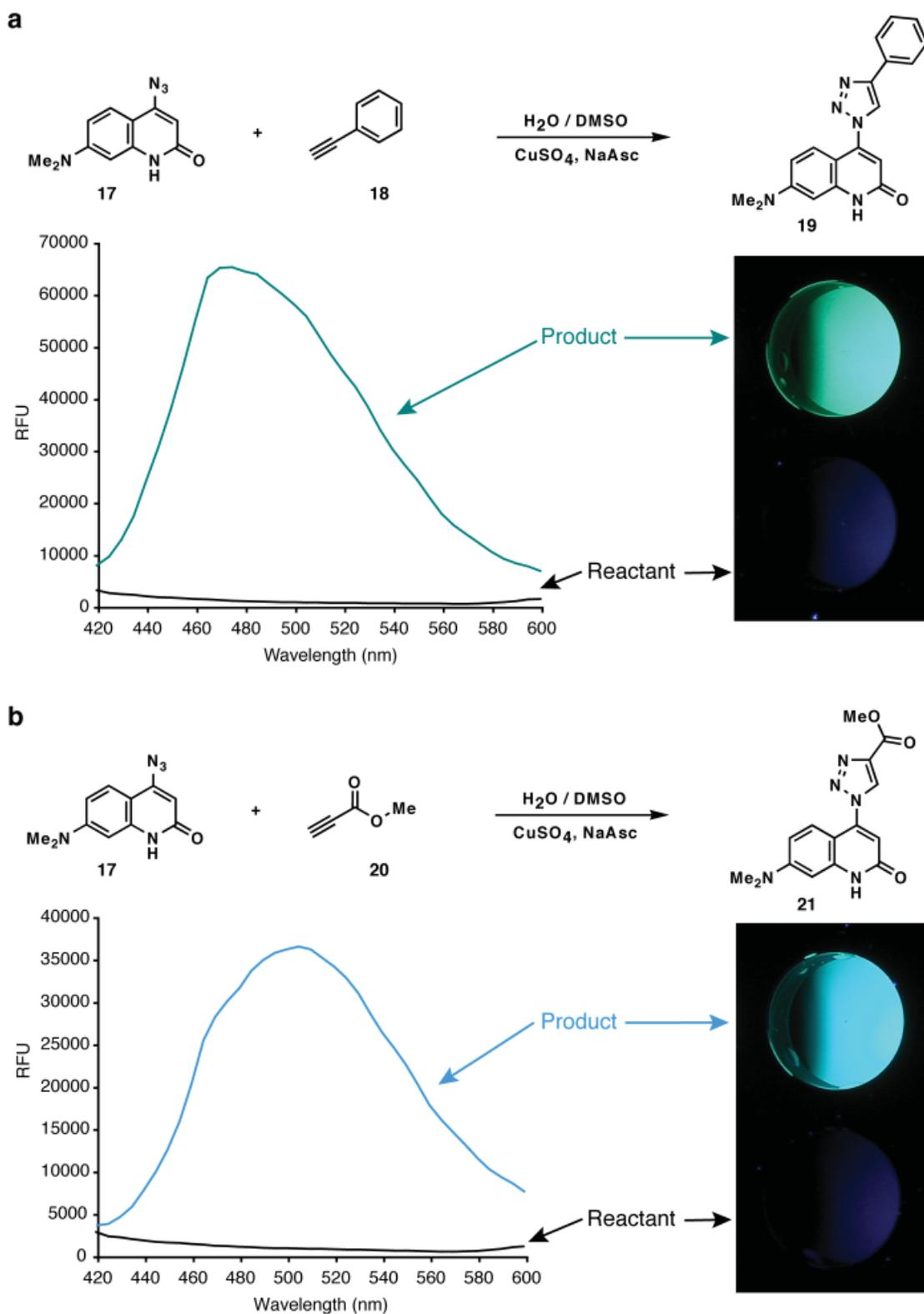
Figure 9.13 Photophysical properties of compound **17** before and after cycloaddition with alkynes **18** and **20**.

Ultrapure RNAse/DNAse free water from USB was used for biological work. All buffer reagents used were molecular biology grade. Buffers were sterilized using a Nalgene 0.2μm cellulose filtration device.

9.4.2 Plasmids

pJDC1 and pJDC2 were constructed using 80mer oligonucleotides purchased from Integrated DNA Technologies. pUC19 plasmid was purchased from Sigma. JM109 Competent Cells (>108 cfu/μg) were purchased from Promega. A Rapid DNA ligation kit was purchased from Roche. Purification was done using a Promega Wizard Plus Midipreps DNA purification system. Sequence analysis was performed by the Sequence Analysis Facility at the Caltech.

9.4.3 Polyamide Synthesis

Polyamide synthesis was performed as previously reported.[13] All polyamides were synthesized using Boc-β-Ala-PAM resin (~0.59 meq/g). The resin was initially swelled in DMF for 5 min in a glass reaction vessel fitted with a glass filter and stopcock. The vessel was drained and the resin washed twice with DCM. Deprotection with 80% TFA:DCM was performed for 20 min while the resin was shaken. Following deprotection, the resin was washed 2 x DCM, 1 x 4:1 DMF:DIEA, and 1 x DMF. Coupling of the Boc-Py-OBt pre-activated ester was performed using 1.8 eq of monomer in ~1 mL of NMP. Coupling of Boc-Im-OH, Boc-PyIm-OH, and other acids was done by first preactivating 1.8 eq of the monomer with 1.7 eq of HBTU, 5.4 eq of DIEA in ~2mL of NMP. The activation mixture was shaken for 20min before being filtered and added to the resin. Couplings were allowed to proceed for 2h except in the case of Im-OH which was allowed to react overnight. After each coupling step the resin was washed 2 x DMF and then 2 x DCM before the next deprotection step. Polyamides were cleaved from resin using 1.5 mL of Dp for 200 mg of resin at 55°C for eight h. Crude products were purified by preparative HPLC.

2: MALDI-TOF-MS calculated [M+H]$^+$: 1348.63, observed 1348.54

3: MALDI-TOF-MS calculated [M+H]$^+$: 1362.65, observed 1362.61

4: MALDI-TOF-MS calculated [M+H]$^+$: 1376.66, observed 1376.73

5: MALDI-TOF-MS calculated [M+H]$^+$: 1289.58, observed 1289.67

6: MALDI-TOF-MS calculated [M+H]$^+$: 1317.61, observed 1317.6

7: MALDI-TOF-MS calculated [M+H]$^+$: 2664.23, observed 2664.88
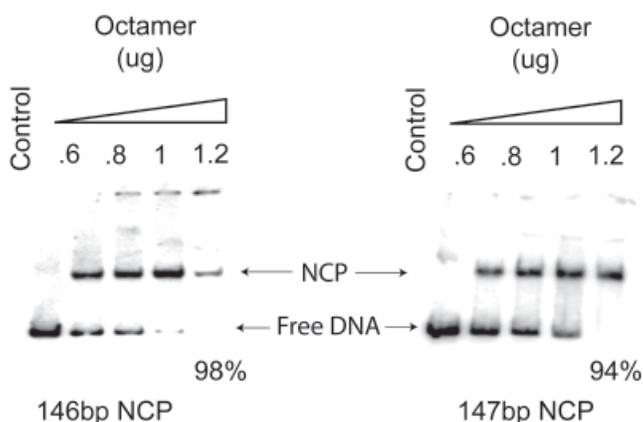
Figure 9.14 Representative gels of the NCP reconstitution. The amount of histone octamer is titrated into a constant amount of the 146 or 147 bp radiolabelled DNA. Non-denaturing polyacrylamide gels are used to determine the yield from reconstitution.

9.4.4 Reconstitution of NCP

The NCP was reconstituted following established protocols. Briefly, 10 μg of the 146 bp DNA was 5' radiolabelled as described above using Polynucleotide Kinase and purified using a Chroma Spin STE 10 column from BD Biosciences. Histone octamer which was obtained from the Luger lab was diluted to 1 μg / μL using 2 M NaCl TE Buffer. 5 μL of the DNA was added to 5 μL of 4M NaCl, 10 mM Tris, 1 mM EDTA solution to make a sample of 10 μL of DNA in 2 M NaCl TE buffer. Five samples were made up and 0, 0.6, 0.8, 1.0, or 1.2 μL of octamer (1 μg / μL) was added to each sample. The total volume was brought to 2 μL using 2M NaCl TE buffer. The samples were incubated for 1 hr before the addition of 12 μL, 6 μL, 6 μL, and 84 μL of dilution buffer (TE) with each addition followed by a 1 h incubation. The sample was heated to 37°C for 2 h before a final addition of 120 μL of dilution buffer to give a final NaCl concentration of .1 M. Reconstituted NCP samples were stored at 4°C. Non-denaturing PAGE was done on a 6% polyacrylamide TBE gel. 1 μL of each sample was diluted to 10 μL with 10 mM Tris, 20 mM NaCl, 10% glycerol, and .1% Igapal. The gel was run at 150 V for 20 min, dried and imaged.

9.4.5 NCP Templated Ligation Reactions

The NCP was reconstituted as described above with the sole difference that non-radiolabelled DNA was used. For each ligation reaction 200 pmol of each polyamide was lyophilized into an eppendorf tube and 40 pmol of the reconstituted NCP was added. Samples containing only the 146 bp DNA and or the histone octamer were prepared as negative controls. The reaction was incubated for 5 h at 37°C before MALDI-TOF analysis.

9.5 Notes and References

1.   Poulin-Kerstien, A. T., and Dervan, P. B. DNA-templated dimerization of hairpin polyamides. *J. Am. Chem. Soc.* **2003,** *125*, 15811-15821.

2. Poulin-Kerstien, A. T. *DNA-Templated Dimerizations of Minor Groove--Binding Polyamides* California Institute of Technology: Pasadena, CA, **2005**

3. Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* **2003,** *326*, 371-380.

4. Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Molecular recognition of the nucleosomal "supergroove". *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101*, 6864-6869.

5. Best, T. P., Edelson, B. S., Nickols, N. G., and Dervan, P. B. Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Acad. Sci. U. S. A.* **2003,** *100*, 12063-12068.

6. Khoukhi, N., Vaultier, M., and Carrie, R. Synthesis and reactivity of methyl-azido butyrates and ethyl-azido valerates and of the corresponding acid chlorides as useful reagents for the aminoalkylation *Tetrahedron.* **1987,** *43*, 1811-1822.

7. Sivakumar, K., Xie, F., Cash, B. M., Long, S., Barnhill, H. N., and Wang, Q. A Fluorogenic 1, 3-Dipolar Cycloaddition Reaction of 3-Azidocoumarins and Acetylenes *Org. Lett.* **2004,** *6*, 4603-4606.

8. Chaurasia, C. S., and Kauffman, J. M. Synthesis and fluorescent properties of a new photostable thiol reagent "BACM." *J. Heterocyc. Chem.* **1990,** *27*, 727-733.

9. *Spartan ES* Wavefunction, Inc: Irvine, CA, **2005**.

10. (a) Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., Defrees, D. J., and Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The. Journal. of. Chemical. Physics.* **1982,** *77*, 3654-3665. (b) Hariharan, P. C., and Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretical. Chemistry. Accounts:. Theory,. Computation,. and. Modeling. (Theoretica. Chimica. Acta).* **1973,** *28*, 213-222.

11. Edayathumangalam, R. S., Weyermann, P., Dervan, P. B., Gottesfeld, J. M., and Luger, K. Nucleosomes in solution exist as a mixture of twist-defect states. *J. Mol. Biol.* **2005,** *345*, 103-114.

12. Nasr, M., Drach, J. C., Smith, S. H., Shipman Jr, C., and Burckhalter, J. H. 7-Aminoquinolines. A novel class of agents active against herpesviruses. *J. Med. Chem.* **1988,** *31*, 1347-1351.

13. Baird, E. E., and Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.