## Supplementary material

### Creation of a type IIS restriction endonuclease with a long recognition sequence

Shaun M. Lippow, Patti M. Aha, Matthew H. Parker, William J. Blake, Brian M. Baynes, and Daša Lipovšek

**Section 1:** Expression screen of I-SceI variants

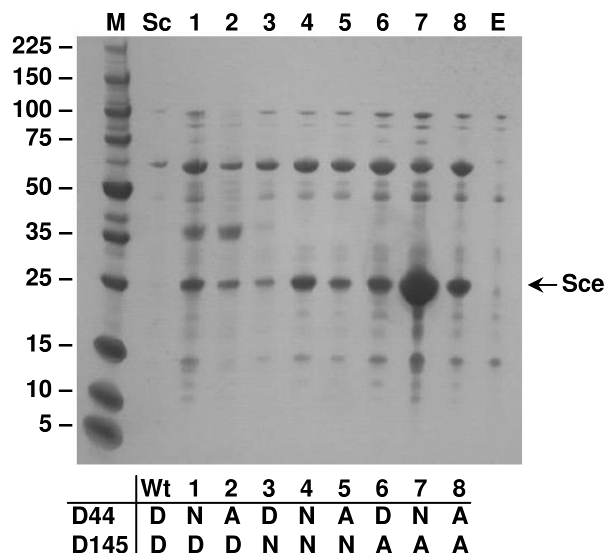


**Figure 1S.** Expression screen of I-SceI variants. M: molecular-weight markers (kD); Sc: wild-type I-SceI; 1–8: I-SceI mutants Sce1–Sce8; E: expression vector without a cloned endonuclease. The residues in positions 44 and 145 in each variant are shown under the gel. Lanes Sc and 1–8 contain 10 μL of the soluble fraction of bacterial lysate expressing each I-SceI variant. Variant Sce7 (lane 7) shows the highest level of soluble expression.

**Section 2:** Modeling and design of hybrid enzymes

Since no structural information is available on FokI bound to and cleaving DNA, we modeled the structural mechanism of the wild-type FokI type IIS endonuclease following the approach previously described by Wah *et al* (1), using available structures of monomeric FokI bound to DNA in an inactive, sequestered conformation (2), dimeric FokI without DNA (1), and BamHI bound to DNA (3). First the DNA-free FokI dimer structure was aligned to a crystal structure of the complex of BamHI and DNA (Figure 2S(a)). In this model, the three catalytic residues of each FokI monomer are in a similar location to their homologous BamHI catalytic residue, with the FokI catalytic residues positioned over DNA-backbone phosphates to produce the four-base overhangs found in FokI digestion products. Next we aligned the DNA backbones from the model of the FokI catalytic domain dimer on DNA (Figure 2S(a)) and from a crystal structure of the FokI monomer bound to DNA (Figure 2S(b)) so that the cleavage domains were positioned at the FokI native cut sites, nine and thirteen bases downstream of the target site (9/13). The cut thirteen bases from the target site is made by the FokI molecule whose DNA-binding domain was modeled to bind the target site, whereas the cut nine bases from the target site is made by the second FokI catalytic domain, which is positioned through nonconvalent interactions between the two catalytic domains. Modeling of the fifteen-residue native linker that connects the C-terminus of the recognition domain to the N-terminus of the closer of the two catalytic domains (Figure 2S(c)) produced several low-energy linker conformations. In the linker models, residues 373 to 377 adopt approximately the same conformation as in the initial sequestered state, residues 378 to 382 adopt a new turn, and residues 383 to 387 connect to and extend the $\alpha$-helix at the N-terminus of the catalytic domain.

An attempt to reposition the FokI cleavage domains even a single base pair away from its 9/13 cleavage site resulted in clashes or large changes in distance between the C-terminal end of the recognition domain and the N-terminus of the FokI cleavage domains. When the cleavage domain was moved one or more base pairs further from the recognition domain, the distance between the two domains became too large to be spanned by the native FokI linker without at least breaking the intramolecular contacts at residues 373 to 377, or stretching into an extended conformation. Conversely, positioning the FokI cleavage domains on DNA closer to the target site resulted in steric clashes between the recognition and catalytic domains (Table 1S(a)).

Chimeric I-SceI/FokI enzymes were designed by first modeling the FokI catalytic domain dimer on DNA at different positions downstream of the I-SceI homing endonuclease. The distance from the C-terminus of I-SceI to the N-terminus of each monomer of the FokI cleavage domain dimer was measured (Table 1S(b)), and linkers were designed to span this distance. We found that the N-terminus of at least one of the FokI catalytic domains is between approximately 20 Å and 35 Å of the C-terminus of I-SceI when the cleavage domain dimer is positioned to cut 1/5, 2/6, or 3/7 nucleotides downstream of the I-SceI target site (Figure 3S). Theory for the end-to-end distance of flexible polypeptides as a function of the number of residues predicts that this distance can be spanned by linkers in the range of 10–20 residues (4,5). A shorter, ten-residue linker would be expected to favor

the 1/5 position, which is closer to the I-SceI target sequence; conversely, a longer, 20-residue linker would be expected to favor the 2/6 and 3/7 positions, which are further from the I-SceI target sequence. We chose four novel linker sequences, two of which are ten amino-acid residues in length, and two of which are 20 residues in length. These designed linkers are rich in glycine, polar, and charged amino-acid residues, and contain few residues with preference for α-helical or β-strand secondary structure (6,7). In addition to the four designed linkers, we tested the linker from wild-type FokI, FokL, which is fifteen residues long.

The coordinates of the protein models are available by email on request.

**Table 1S.** Modeling the FokI catalytic dimer at different positions on DNA relative to the FokI and I-SceI target domains.

| Position of catalytic domains on DNA[a] | # catalytic-domain residues that clash with the target domain | Distance from C-terminus of target domain to N-terminus of each catalytic domain (Å) | |
|---|---|---|---|
| | | Upstream catalytic domain | Downstream catalytic domain |
| (a) Downstream of FokI target domain (native FokI) | | | |
| 7/11 | 20 | 70 | 39 |
| 8/12 | 3 | 70 | 23 |
| 9/13[b] | 0 | 67 | 24 |
| 10/14 | 0 | 59 | 33 |
| 11/15 | 0 | 53 | 48 |
| (b) Downstream of I-SceI target domain (chimeric endonuclease) | | | |
| 0/4 | 3 | 45 | 15 |
| 1/5[c] | 0 | 41 | 22 |
| 2/6[c] | 0 | 34 | 34 |
| 3/7[c] | 0 | 28-36 | 38-44 |
| 4/8 | 0 | 37 | 46 |
| 5/9 | 0 | 43 | 50 |

[a] Defined by predicted number of nucleotides between 3' end of target site and site of DNA cleavage.
[b] Model consistent with observed wild-type FokI cleavage. The fifteen-residue native linker connects the C-terminus of the DNA-bound target domain to the N-terminus of the downstream catalytic domain, 24 Å away.
[c] Models used to design linkers for chimeric I-SceI/FokI endonucleases, based on the absence of catalytic-domain clashes and at least one distance between domain termini shorter than 35 Å.

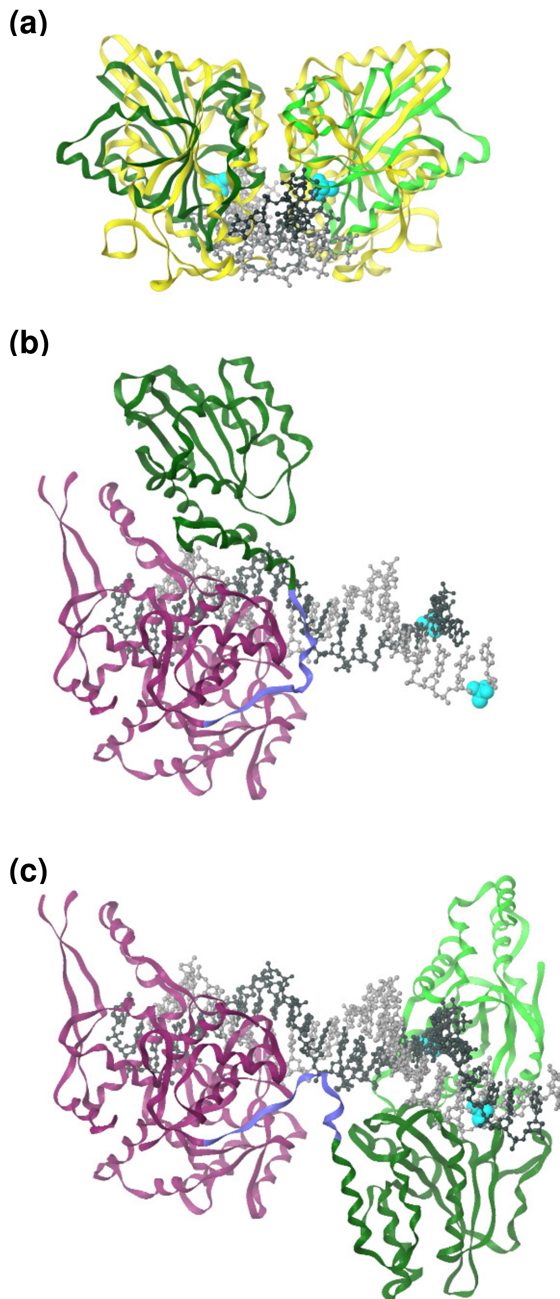


**Figure 2S.** Molecular modeling of the wild-type FokI enzyme. (a) Alignment of BamHI (yellow, 1BHM (3)) to FokI catalytic domains (green, 2FOK (1)), with DNA in gray and BamHI cleavage sites in cyan. (b) Crystal structure of FokI (1FOK (2)), with binding domain in magenta, linker in purple, catalytic domain in green, DNA in gray, and DNA cleavage sites in cyan. (c) Model of active state of FokI. Of the second FokI monomer, only the catalytic domain is shown.

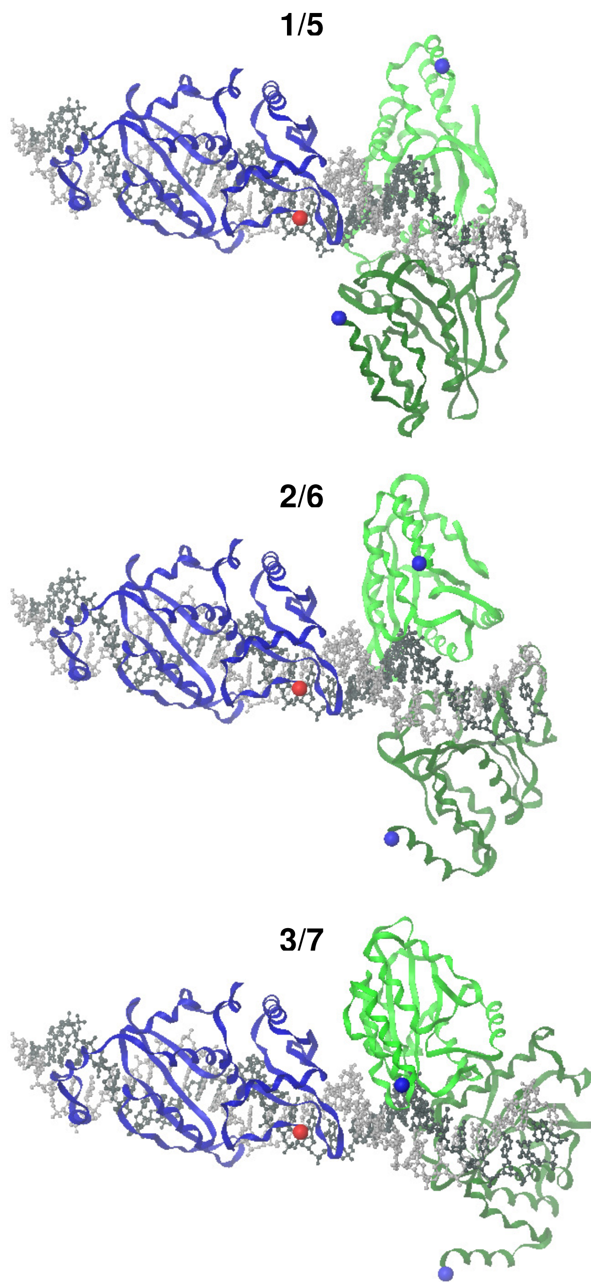


**Figure 3S.** Molecular modeling of hybrid I-SceI/FokI enzymes. I-SceI is shown in blue, with a red sphere at the C-terminus, and the FokI catalytic dimer is shown in green, with blue spheres at both N-termini. The recognition domain of the second monomer is not shown. Each panel shows the catalytic dimer positioned to cleave DNA at a different number of base pairs downstream of the I-SceI recognition site as indicated. Peptide linkers of 10–20 residues were designed to connect the C-terminus of the I-SceI–derived DNA-recognition domain to the N-terminus of one of the FokI catalytic domains.

**Section 3:** Sequence of plasmids used to test assembly of DNA fragments generated by chimeric endonucleases

>pCK
agtgccacctgacgtttaattaatgcagaaataaaaaggcctgcgattaccagcaggcctgttattagct
cagtagacgtcttacgccccgccctgccactcatcgcagtactgttgtaattcattaagcattctgccga
catggaagccatcacaaacggcatgatgaacctgaatcgccagcggcatcagcaccttgtcgccttgcgt
ataatatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactg
gtgaaactcacccagggattggctgatacgaaaaacatattctcaataaaccctttagggaaataggcca
ggttttcaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattc
actccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactatcccat
atcaccagctcaccgtctttcattgccatacggaattccggatgagcattcatcaggcgggcaagaatgt
gaataaaggccggataaaacttgtgcttatttttctttacggtctttaaaaaggccgtaatatccagctg
aacggtctggttataggtacattgagcaactgactgaaatgcctcaaaatgttctttacgatgccattgg
gatatatcaacggtggtatatccagtgatttttttgttcaagagaccgcggccgccgtctcttgagccat
attcaacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctgatttatatgggtata
aatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgattgtatgggaagcccgatgcgcc
agagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaac
tggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttac
tcaccactgcgatccccgggaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatat
tgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagc
gatcgcgtatttcgcctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttg
atgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaactgttgccattctcacc
ggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggt
tgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcg
gtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataa
attgcagtttcatttgatgctcgatgagtttttctaagctcttccgctagctttcaaataaaacgaaagg
ctcagtcgaaagactgggcctttcgttttaatctggtttaaacgcttcctcgctcactgactcgctgcgc
tcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcag
gggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgtt
gctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtgg
cgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttc
cgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctc
acgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgtt
cagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgc
cactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaa
gtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttacc
ttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgttt
gcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctga
cgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctag
atccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagtt
accaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgact
ccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcga
gaaccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtg
gtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgcc
agttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatg
gcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcgg
ttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggc
agcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaacc
aagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccg
cgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggat
cttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttact
ttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacac
ggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcat
gagcggatacatatttgaatgtatttagaaaataaacaaataggggttccgcgcacatttccccgaaa

```
>pTS
agtgccacctgacgtttaattaatgcagaaataaaaaggcctgcgattaccagcaggcctgttattagct
cagtagacgtcttaggtcgaggtggcccggctccatgcaccgcgacgcaacgcggggaggcagacaaggt
atagggcggcgcctacaatccatgccaacccgttccatgtgctcgccgaggcggcataaatcgccgtgac
gatcagcggtccagtgatcgaagttaggctggtaagagccgcgagcgatccttgaagctgtccctgatgg
tcgtcatctacctggcgggacagcatggcctgcaacgcgggcatcccgatgccgccggaagcgagaagaa
tcataatggggaaggccatccagcctcgcgtcgcgaacgccagcaagacgtagcccagcgcgtcggccgc
catgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcg
agggcgtgcaagattccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcct
cgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaagaatacagtcataag
tgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatc
ggtcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagca
ccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccac
catacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcg
gcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagagga
tccacaggacgggtgtggtcgccatgatcgcgtagtcgatagtggctccaagtagcgaagcgagcaggac
tgggcggcggccaaagcggtcggacagtgctccgagaacgggtgcgcatagaaattgcatcaacgcatat
agcgctagcagcacgccatagtgactggcgatgctgtcggaatggacgatatcccgcaagaggcccggca
gtaccggcataaccaagcctatgcctacagcatccagggtgacggtgccgaggatgacgatgagcgcatt
gttagatttcaagagaccgcggccgccgtctcttgagggaagcggtgatcgccgaagtatcgactcaact
atcagaggtagttggcgtcatcgagcgccatctcgaaccgacgttgctggccgtacatttgtacggctcc
gcagtggatggcggcctgaagccacacagtgatattgatttgctggttacggtgaccgtaaggcttgatg
aaacaacgcggcgagctttgatcaacgaccttttggaaacttcggcttcccctggagagagcgagattct
ccgcgctgtagaagtcaccattgttgtgcacgacgacatcattccgtggcgttatccagctaagcgcgaa
ctgcaatttggagaatggcagcgcaatgacattcttgctggtatcttcgagccagccacgatcgacattg
atctggctatcttgctgacaaaagcaagagaacatagcgttgccttggtaggtccagcggcggaggaact
ctttgatccggttcctgaacaggatctatttgaggcgctaaatgaaaccttaacgctatggaactcgccg
cccgactgggctggcgatgagcgaaatgtagtgcttacgttgtcccgcatttggtacagcgcagtaaccg
gcaaaatcgcgccgaaggatgtcgctgccgactgggcaatggagcgcctgccggcccagtatcagcccgt
catacttgaagctagacaggcttatcttggacaagaagaagatcgcttggcctcgcgcgcagatcagttg
gaagaatttgtccactacgtgaaaggcgagatcaccaaggtagtcggcaaataagctcttccgctagctt
tcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttaatctggtttaaacgcttcctcgc
tcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacg
gttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaac
cgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgac
gctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccct
cgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtg
gcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtg
tgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggt
aagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggt
gctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctc
tgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtag
cggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatc
ttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaa
aaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagta
aacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttca
tccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtg
ctgcaatgataccgcgagaaccacgctcaccggctccagatttatcagcaataaaccagccagccggaag
ggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagct
agagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcac
gctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccat
gttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgtta
tcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtga
ctggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtc
aatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcgggg
cgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgat
```

```
cttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa
gggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttat
cagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgc
gcacatttccccgaaa
```

## **Section 4:** References


1. Wah, D.A., Bitinaite, J., Schildkraut, I. and Aggarwal, A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc. Natl. Acad. Sci. USA*, **95**, 10564-10569.
2. Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97-100.
3. Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1995) Structure of Bam HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science*, **269**, 656-663.
4. Zhou, H.X. (2003) Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen. *J. Mol. Biol.*, **329**, 1-8.
5. Zhou, H.X. (2004) Polymer models of protein stability, folding, and interactions. *Biochem.*, **43**, 2141-2154.
6. Williams, R.W., Chang, A., Juretic, D. and Loughran, S. (1987) Secondary structure predictions and medium range interactions. *Biochim. Biophys. Acta*, **916**, 200-204.
7. Wilmot, C.M. and Thornton, J.M. (1988) Analysis and prediction of the different types of beta-turn in proteins. *J. Mol. Biol.*, **203**, 221-232.