Systematic single base-pair substitution analysis of DNA binding by the cAMP receptor protein in cyanobacterium *Synechocystis* sp. PCC 6803

Katsumi Omagari[a], Hidehisa Yoshimura[a], Mitsunori Takano[a], Dongyun Hao[b], Masayuki Ohmori[a], Akinori Sarai[c], Akira Suyama[a,*]

[a] *Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Komaba, Meguro, Tokyo 153-8902, Japan*
[b] *Key Laboratory for Molecular Enzymology and Engineering of the Ministry of Education (Jilin University), Changchun 130021, PR China*
[c] *Department of Biochemical Engineering and Science, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan*

Received 29 January 2004; revised 26 February 2004; accepted 26 February 2004

First published online 12 March 2004

Edited by Thomas L. James

Abstract **The cAMP receptor protein SYCRP1 in cyanobacterium *Synechocystis* sp. PCC 6803 is a regulatory protein that binds to the consensus DNA sequence (5′-AAATGTGATCTAGATCACATTT-3′) for the cAMP receptor protein CRP in *Escherichia coli*. Here we examined the effects of systematic single base-pair substitutions at positions 4–8 (TGTGA) of the consensus sequence on the specific binding of SYCRP1. The consensus sequence exhibited the highest affinity, and the effects of base-pair substitutions at positions 5 and 7 were the most deleterious. The result is similar to that previously reported for CRP, whereas there were differences between SYCRP1 and CRP in the rank order of affinity for each substitution.**

Key words: Protein–DNA interaction; Regulatory protein; Binding free energy change; Systematic base-pair substitution; cAMP receptor protein; Cyanobacterium

1. Introduction

cAMP acts as an intracellular signaling molecule in both prokaryotic and eukaryotic cells. Cells regulate gene expression by changing the intracellular concentration of cAMP in response to environmental changes [1]. Some prokaryotic cells have a regulatory protein that binds to specific DNA sequences in the presence of cAMP, turning on and off hundreds of genes. Recently, it has been found that cyanobacterium *Synechocystis* sp. PCC 6803 also possesses this type of cAMP receptor protein, which is referred to as SYCRP1 [2]. The amino acid sequence of SYCRP1 shows only 23% similarity to that of *Escherichia coli* cAMP receptor protein (CRP), which is a well-studied cAMP receptor protein the three-dimensional structure of which has already been determined [3,4]. Despite the low amino acid sequence similarity, SYCRP1 has been shown to bind specifically to the 22 bp consensus DNA sequence (5′-$A_1A_2A_3T_4G_5T_6G_7A_8T_9C_{10}T_{11}$-$A_{12}G_{13}A_{14}T_{15}C_{16}A_{17}C_{18}A_{19}T_{20}T_{21}T_{22}$-3′ [5]) for CRP in the presence of cAMP [3]. This can be understood by considering that the three-dimensional structure of SYCRP1 is predicted to be very similar to that of CRP [6], and that three key amino acids, Arg-196, Glu-197, and Arg-201 in the helix-turn-helix motif of the DNA-binding domain in SYCRP1, are located at almost the same positions as found in CRP [4]. It is therefore likely that the interaction between SYCRP1 and DNA is similar to that found in the CRP–DNA complex. However, it is not clear which base-pairs of the consensus DNA sequence contribute to the specific binding between SYCRP1 and DNA.

In this study, we addressed the binding mechanism between SYCRP1 and DNA by identifying which base-pairs in the consensus DNA sequence play crucial roles in binding specificity. For this purpose, we measured the binding free energy changes between SYCRP1 and DNA by conducting systematic single base-pair substitution experiments [7–11]. We synthesized the consensus DNA sequence (wild-type) for CRP and point-mutated sequences consisting of all possible single base-pair substitutions at positions 4–8 ($T_4G_5T_6G_7A_8$), which is the highly conserved sequence for CRP binding, of the consensus DNA sequence. We found that the consensus (wild-type) sequence showed the highest affinity, and that substitutions at G_5 and G_7 resulted in significant reduction of binding free energy.

2. Materials and methods

2.1. Preparation of SYCRP1

SYCRP1 used in this study was prepared by the method established by Yoshimura et al. [2]. The purified SYCRP1 was suspended in 50 mM Tris–HCl (pH 8.0), 200 mM NaCl, and 50% glycerol, then stored at −20°C. The concentrations of SYCRP1 were measured using a protein assay kit II (Bio-Rad), and additional confirmation was obtained by the method proposed by Gill and von Hippel [12].

2.2. Preparation of DNA

We prepared 16 fragments of 40 bp double-stranded DNA with a single protruding base G at the 5′ ends (Fig. 1). One of the DNA fragments, ICAP, which contains the 22 bp consensus sequence for CRP, was used as the wild-type sequence in this study. The sequence of ICAP is 5′-(G)CAACGCAATAAATGTGATCTAGATCACATTTTAGGCACCC-3′, where the 5′-protruding base is in parentheses. The remaining fragments were prepared by systematic base-pair substitution within the underlined portion (positions 4–8) in the ICAP sequence. In addition, we also prepared another 40 bp double-stranded DNA fragment, slr1667, which was shown to be regulated by SYCRP1 [6]. The sequence of slr1667 is 5′-(G)ATACACAACAGTTGTGATCTGGGTCACAACCATTGAGTGA-3′, where the underlined bases and the base in parentheses indicate the bases identical with those of ICAP and the 5′-protruding base, respectively.

*Corresponding author. Fax: (81)-3-5454-6528.
E-mail address: suyama@dna.c.u-tokyo.ac.jp (A. Suyama).

Abbreviations: SYCRP1, cAMP receptor protein in cyanobacterium *Synechocystis* sp. PCC6803; CRP, cAMP receptor protein

0014-5793/04/$30.00
doi:10.1016/S0014-5793(04)00248-0

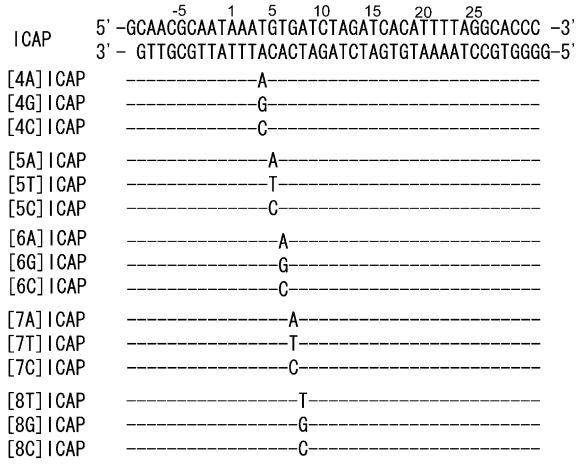

Fig. 1. DNA sequences used in this study. The base numbering follows that used in [11]. ICAP is the consensus DNA for CRP. Positions 4–8 in ICAP were subjected to systematic single base-pair substitution. All the possible DNA sequences with single base-pair substitution are shown.

On the other strand, the 5′-protruding base was A. The highly conserved bases of CRP binding are printed in boldface.

The DNA used in this study was commercially synthesized and purified by high performance liquid chromatography. Each oligonucleotide was suspended in 50 mM Tris–HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, and annealed in the presence of its complementary strand so as to obtain a double-stranded form. The resulting DNA fragments were labeled with [γ-^{32}P]ATP (Amersham) using T4 polynucleotide kinase (Toyobo). DNA fragments thus prepared were finally purified through a Sephadex G-50 column (Pharmacia). Then the purified DNA fragments were used in the titration experiments.

2.3. Binding titration experiment and determination of $\Delta\Delta$G

To calculate the binding free energy ΔG, we must determine the equilibrium constant for SYCRP1–DNA binding. The equilibrium dissociation constant K_d for SYCRP1–DNA binding was measured by quantitative electrophoresis mobility shift assay (EMSA) [2,10,13]. The labeled DNA sample, at a concentration 10–1000-fold lower than the K_d value, was mixed with a gradient concentration of SYCRP1 in 30 μl of binding buffer (50 mM Tris–HCl (pH 7.5), 60 mM NaCl, 1 mM EDTA, 8.3% glycerol, 0.1 mg/ml bovine serum albumin) with a final concentration of 20 μM cAMP. After incubation at about 22°C for 30 min to reach equilibrium, the samples were quickly loaded onto a 10% non-denaturing polyacrylamide gel. Electrophoresis was performed for 30–45 min at 400 V in 0.25×TBE buffer with 20 μM cAMP. The results were not affected by the electrophoresis time. The resulting gel was dried and visualized in a Fujix BAS1000 system (Fuji film). The photo-densities of the SYCRP1–DNA complex and the free DNA bands were determined.

We assumed that the binding reaction obeys the following reaction scheme:

$$\mathrm{D} + \mathrm{S} \rightleftharpoons \mathrm{DS} \tag{1}$$

where D and S denote DNA and SYCRP1 with a cAMP molecule, respectively, and DS represents their complex. The equilibrium dissociation constant for this reaction is given by:

$$K_d = \frac{[D][S]}{[DS]} \tag{2}$$

Simple transformation of Eq. 2 and the assumption of $[D] \ll K_d$ give us the following form:

$$y = a + \frac{b}{1 + x/K_d} \tag{3}$$

where x denotes the SYCRP1 concentration added to the reaction buffer and y denotes the fraction of the free DNA, i.e. [D]/([D]+[DS]). The parameters a and b are mere constants to adjust the baseline. Eq. 3 was used to fit the titration experimental data, and K_d was obtained as a result of fitting. The fitting was done by non-linear regression based on the Marquardt–Levenberg algorithm as implemented in IGOR Pro 3.1 (WaveMetrics). The binding free energy change $\Delta\Delta G$ was calculated as:

$$\Delta\Delta G = \Delta G_{\mathrm{mutant}} - \Delta G_{\mathrm{ICAP}} = RT \ln \frac{K_d^{\mathrm{mutant}}}{K_d^{\mathrm{ICAP}}} \tag{4}$$

where R is the gas constant and T is the absolute temperature. Positive $\Delta\Delta G$ indicates reduction of binding affinity. Note that a 10-fold decrease in binding affinity corresponds to $\Delta\Delta G \approx 1.3$ kcal/mol (at 22°C).

3. Results

Fig. 2 shows typical gel images for SYCRP1 binding to the consensus sequence ICAP and the point-mutated ICAP sequences in Fig. 1. It is shown that the fraction of the SYCRP1–DNA complex became larger and that of the free double-stranded DNA became smaller as the concentration of the added SYCRP1 increased. From the photo-density for the SYCRP1–DNA complex band and that for the free double-stranded DNA band, we calculated the fraction of free double-stranded DNA as a function of added SYCRP1 concentration. In Fig. 3, the titration curves thus obtained for ICAP and for a point-mutated ICAP ([7A]ICAP) are presented. The dissociation constant K_d can be roughly estimated from the concentration of the added SYCRP1 at which the fraction of the free DNA becomes one-half. Fig. 3 shows that the K_d value for the point-mutated ICAP ([7A]ICAP) was 20 times larger than that for the consensus ICAP, indicating that introduction of a single base-pair substitution causes a 1.8 kcal/mol increase in binding free energy. The titration experimental

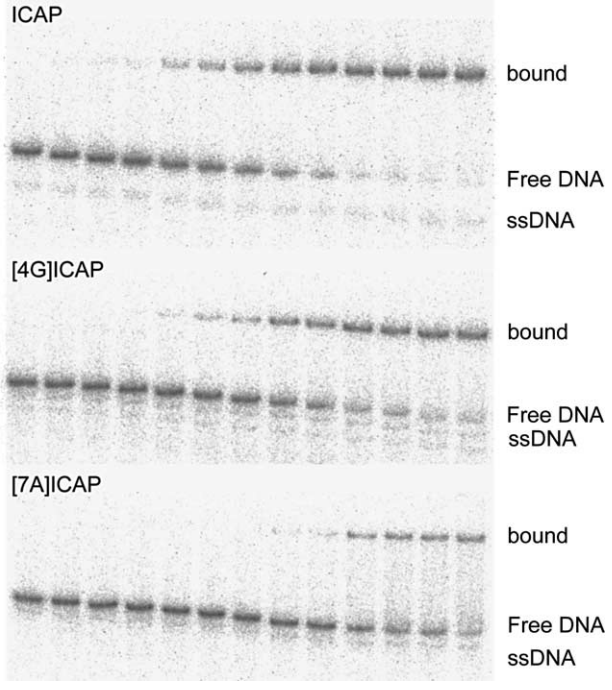

Fig. 2. Quantitative EMSA for SYCRP1 binding to DNA. Typical autoradiograms of gels are shown for ICAP, [4G]ICAP, and [7A]ICAP. From the left to the right lanes, final protein concentrations are 0 M, 6.6 pM, 13 pM, 33 pM, 66 pM, 130 pM, 330 pM, 0.66 nM, 1.3 nM, 3.3 nM, 6.6 nM, 13 nM, and 33 nM, respectively.

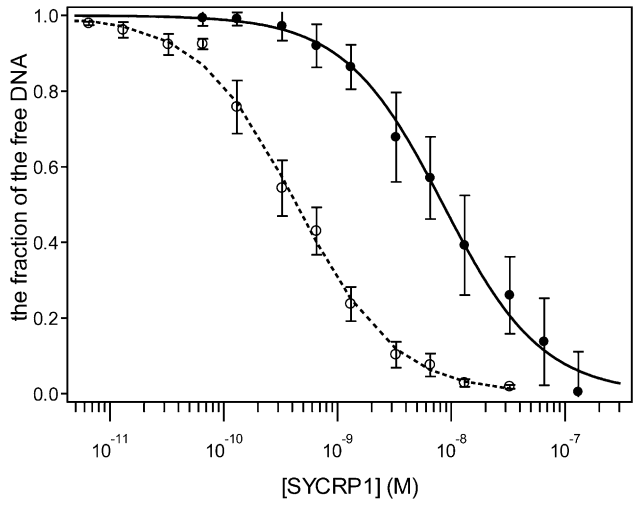

Fig. 3. Titration curves for the SYCRP1–ICAP binding (○) and the SYCRP1–[7A]ICAP binding (●). The fraction of free DNA is shown as a function of the concentration of SYCRP1. Error bars represent standard errors calculated from three independent experiments. The solid and the broken lines are the best fitting curves obtained by non-linear regression.

data were well fitted to the theoretical curve derived from the assumption of the simple binding reaction scheme (Eqs. 1–3). Thus, all K_d values were obtained by fitting the titration experimental data to Eq. 3.

By using the obtained K_d values, we calculated the binding free energy changes caused by the systematic single base-pair substitution. The $\Delta\Delta G$ values for the respective single base-pair substitutions are summarized in Fig. 4. A positive $\Delta\Delta G$ value means that the binding affinity is reduced by the base-pair substitution. Fig. 4 shows that all of the substitutions

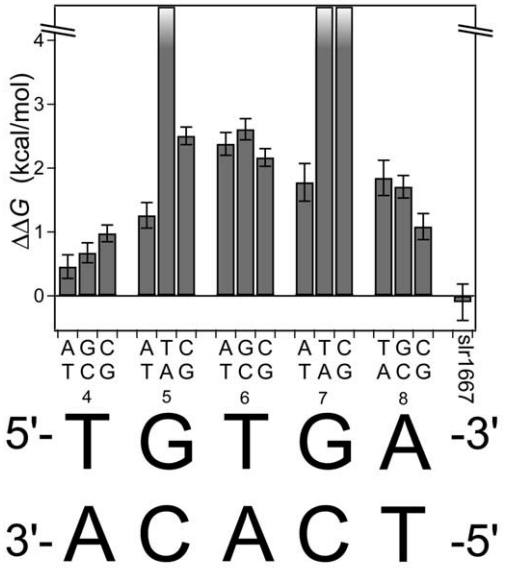

Fig. 4. Effect of single base-pair substitutions on binding free energy change $\Delta\Delta G$. The sequence shown at the bottom is that of ICAP. Each bar shows the $\Delta\Delta G$ corresponding to the base substitution indicated at the bottom. K_d values were determined by quantitative EMSA and transformed into $\Delta\Delta G$ values as described in Section 2. Error bars are the standard errors calculated from three independent experiments.

resulted in an increase of binding free energy, indicating that ICAP possesses the highest binding affinity with SYCRP1. Substitutions at G_5 and G_7 strongly affected the binding affinity, resulting in significant increases (of more than 4 kcal/mol for T or C substitution) in binding free energy. This indicates that G_5 and G_7 of the consensus sequence play crucial roles in binding specificity between SYCRP1 and DNA. In contrast to the substitutions of G→T/C at positions 5 and 7, other substitutions showed milder but non-negligible changes (2–3 kcal/mol or less) in binding free energy. In addition to single base-pair substitutions of the base-pairs in the TGTGA motif, we also examined the effect of substitution outside the TGTGA motif on the binding affinity by using the actual promoter sequence of the slr1667 gene of cyanobacterium [6], which contains the TGTGA motif and is known to be regulated by SYCRP1. Fig. 4 for slr1667 shows that substitution outside the TGTGA motif exerted a negligible influence on the binding free energy.

4. Discussion

We clarified that G:C base-pairs at positions 5 and 7 within the consensus sequence, 5′-AAATGTGATCTAGATCACA-TTT-3′, exhibited crucial contributions to the binding specificity between DNA and SYCRP1, and that other base-pairs within the TGTGA motif also contributed to some extent to the specificity. Given that substitution outside the TGTGA motif had little effect on the binding free energy, the highly conserved base-pairs, TGTGA, in the consensus DNA sequence for CRP appear to be sufficient for the specific binding of SYCRP1.

It is worthwhile to compare our results of SYCRP1 of cyanobacterium with those of other homologous proteins of different organisms. As mentioned above, CRP in *E. coli* was previously found to exhibit high affinity with the consensus ICAP sequence used in this study [14]. It is interesting that the base-pairs involved in specificity for CRP, that is, G:C pairs at positions 5 and 7 [11,15], are the same as those found in our study. Although SYCRP1 has only low amino acid sequence similarity to CRP in total (23%), it shows a high local sequence similarity to CRP in the domain corresponding to the DNA-binding domain of CRP (Fig. 5). According to the crystal three-dimensional structure, this domain of CRP forms a helix-turn-helix motif, and one of the helices fits into the major groove of DNA [4]. In the helix, three amino acids (Arg180, Glu181, and Arg185) are shown to interact directly with G:C base-pairs at positions 5 and 7 through hydrogen bonding. With regard to SYCRP1, although its three-dimensional structure has not yet been determined experimentally,

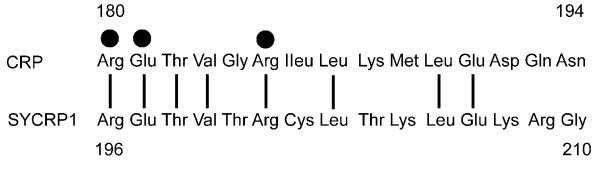

Fig. 5. The amino acid sequence in the recognition helix of CRP and the corresponding sequence for SYCRP1. The amino acid residues from 180 to 194 in CRP appear in the recognition helix of CRP. The closed circles (●) represent the key amino acids that are found to form hydrogen bonds with base-pairs at positions 5 and 7 [4]. Identical amino acids between CRP and SYCRP1 are indicated by vertical lines.

the homology-modeling method [16] predicted that its three-dimensional structure is very similar to that of CRP, with the above-mentioned three functional amino acid residues positioned almost as they are in CRP [6]. Therefore, for the SYCRP1–DNA binding as well, it is very likely that G:C pairs at positions 5 and 7 are involved in direct contact with the conserved three key amino acid residues of SYCRP1 through hydrogen bonding. The free energy increase of more than 4 kcal/mol caused by substitutions at positions 5 and 7 (Fig. 4) is consistent with the energy for more than two hydrogen bonds.

The catabolite gene activator protein-like protein in *Xanthomonas campestris* (CLP) shows 45% amino acid sequence similarity to CRP and also has the three key amino acid residues [17]. The base-pairs involved in the specificity for CLP–DNA binding are the same as those in the SYCRP1 and CRP cases, i.e. G:C pairs at positions 5 and 7. Considering these facts on SYCRP1, CRP, and CLP, it can be speculated that the local interactions between the G:C pairs and the three conserved functional amino acid residues are of importance in the binding specificity between DNA and CRP-like proteins.

Although the overall trend of binding free energy changes due to systematic single base-pair substitution is similar to that reported in CRP, there were differences between SYCRP1 and CRP in the rank order of binding free energy changes for each substitution. With regard to the substitution of the G:C pair at position 5, $\Delta\Delta G$ is ranked so that A:T < C:G < T:A for SYCRP1 (Fig. 4) and T:A < A:T < C:G for CRP [11]. In SYCRP1, in contrast with CRP, the A:T substitution is preferred to the T:A substitution. With regard to the substitution of the G:C pair at position 7, the rank order of $\Delta\Delta G$ for SYCRP1 is A:T ≪ T:A ≈ C:G (Fig. 4), and that for CRP is T:A ≪ A:T ≈ C:G [11]. Again in SYCRP1 and in contrast with CRP the A:T substitution is preferred to the T:A substitution.

Without detailed three-dimensional structure knowledge of the SYCRP1–DNA complex, we cannot address the reason why these differences in rank order of $\Delta\Delta G$ appeared. However, one possible explanation is that the differences in rank order of $\Delta\Delta G$ may be caused by the local interactions with non-conserved amino acids. The observed differences in rank order may also suggest that non-local or indirect interactions, including an allosteric effect like that found in CRP [18,19], play a role in modulating binding affinity. Although examination of these possibilities is beyond the scope of the present study, it is essential to accumulate further fundamental knowledge of the effect of amino acid substitutions on binding affinity and to determine the three-dimensional structure of both the isolated protein and the complex with DNA.

Acknowledgements: This work was supported in part by a grant-in-aid from The 21st Century COE (Center of Excellence) program (Research Center for Integrated Science) of the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

References

[1] Botsford, J.L. and Harman, J.G. (1992) Microbiol. Rev. 56, 100–122.
[2] Yoshimura, H., Hisabori, T., Yanagisawa, S. and Ohmori, M. (2000) J. Biol. Chem. 275, 6241–6245.
[3] Weber, I.T. and Steitz, T.A. (1987) J. Mol. Biol. 198, 311–326.
[4] Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y.W., Ebright, R.E. and Berman, H.M. (1996) J. Mol. Biol. 260, 395–408.
[5] Berg, O.G. and von Hippel, P.H. (1988) J. Mol. Biol. 200, 709–723.
[6] Yoshimura, H., Yanagisawa, S., Kanehisa, M. and Ohmori, M. (2002) Mol. Microbiol. 43, 843–853.
[7] Takeda, Y., Sarai, A. and Rivera, V.M. (1989) Proc. Natl. Acad. Sci. USA 86, 439–443.
[8] Sarai, A. and Takeda, Y. (1989) Proc. Natl. Acad. Sci. USA 86, 6513–6517.
[9] Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S. and Sarai, A. (1993) Proc. Natl. Acad. Sci. USA 90, 9320–9324.
[10] Hao, D., Ohme-Takagi, M. and Sarai, A. (1998) J. Biol. Chem. 273, 26857–26861.
[11] Gunasekera, A., Ebright, Y.W. and Ebright, R.H. (1992) J. Biol. Chem. 267, 14713–14720.
[12] Gill, S.C. and von Hippel, P.H. (1989) Anal. Biochem. 182, 319–326.
[13] Hao, D., Yamasaki, K., Sarai, A. and Ohme-Takagi, M. (2002) Biochemistry 41, 4202–4208.
[14] Ebright, R.H., Ebright, R.W. and Gunasekera, A. (1989) Nucleic Acids Res. 17, 10295–10305.
[15] Jansen, C., Gronenborn, A.M. and Clore, G.M. (1987) Biochem. J. 246, 227–232.
[16] Ogata, K. and Umeyama, H. (2000) J. Mol. Graph. Model. 18, 258–272, 305–306.
[17] Dong, Q. and Ebright, R.H. (1992) J. Bacteriol. 174, 5457–5461.
[18] Leu, S.F., Baker, C.H., Lee, E.J. and Harman, J.G. (1999) Biochemistry 38, 6222–6230.
[19] Lin, S.H., Kovac, L., Chin, A.J., Chin, C.C. and Lee, J.C. (2002) Biochemistry 41, 2946–2955.