882–887 Nucleic Acids Research, 1999, Vol. 27, No. 3

Using sequence logos and information analysis of Lrp DNA binding sites to investigate discrepancies between natural selection and SELEX

Ryan K. Shultzaberger[1,2,+] and Thomas D. Schneider[2,*]

[1]Catoctin High School, 14745 Sabillasville Road, Thurmont, MD 21788, USA and [2]Laboratory of Mathematical Biology, National Cancer Institute, Frederick Cancer Research and Development Center, PO Box B, Building 469, Room 144, Frederick, MD 21702-1201, USA

Received August 26, 1998; Revised and Accepted December 1, 1998

ABSTRACT

***In vitro* experiments that characterize DNA–protein interactions by artificial selection, such as SELEX, are often performed with the assumption that the experimental conditions are equivalent to natural ones. To test whether SELEX gives natural results, we compared sequence logos composed from naturally occurring leucine-responsive regulatory protein (Lrp) binding sites with those composed from SELEX-generated binding sites. The sequence logos were significantly different, indicating that the binding conditions are disparate. A likely explanation is that the SELEX experiment selected for a dimeric or trimeric Lrp complex bound to DNA. In contrast, natural sites appear to be bound by a monomer. This discrepancy suggests that *in vitro* selections do not necessarily give binding site sets comparable with the natural binding sites.**

INTRODUCTION

Genetic control is exerted when proteins bind to specific nucleic acid sequences. Traditionally, these sequences have been collected from the naturally evolved sites. More recently, protein-binding motifs have been characterized by using *in vitro* selection procedures. It is often assumed that *in vitro* results accurately reflect natural binding sites, but a quantitative comparison of the two approaches has usually been lacking. In this paper we make this comparison for the leucine-responsive regulatory protein (Lrp).

Lrp is a pleiotropic DNA-binding protein in *Escherichia coli* and *Salmonella typhimurium* that consists of two 18.8 kDa subunits (1), and that forms a homodimer in solution (2). Lrp binds to multiple sites in a number of operons, including *dad*, *fanABC*, *papBA* and *ilvIH* (3–5). Leucine can invoke either positive or negative transcriptional control by Lrp (1,6).

Cui *et al*. investigated Lrp by using the SELEX (systematic evolution of ligands by exponential enrichment) procedure (7), an *in vitro* method that is used to identify binding motifs. In the SELEX procedure, a specific protein is used to select binding sequences from random synthetic sequences (8). Since its introduction, the SELEX technique has been used to study a variety of systems (9,10).

Since Lrp has many natural binding sites, a reasonably accurate model for *in vivo* binding sequences can be created and compared with sites produced by SELEX. Based on Claude Shannon's information theory (11,12), molecular information theory (13,14) is a mathematical approach to explaining molecular interactions. Using information theory, we constructed two separate models of Lrp binding sequences for comparison. These quantitative models, called sequence logos (15), graphically represent Lrp binding in both the natural and synthetic environments. Comparison of the models allowed us to test whether the sites selected *in vitro* had evolved to simulate natural binding sites.

MATERIALS AND METHODS

Twenty-seven Lrp binding sites were aligned for analysis of Lrp binding patterns (Fig. 1). Only sites with supporting experimental data (footprint, mutational analysis, deletions) were used. Seventeen of the 27 sites are on the *E.coli* chromosome and the remaining 10 are found on four different plasmids. The *lrp*, *gltBDF*, *leuABCD*, *oppA*, *pnt*, *sdaA*, *glyA*, *livJ*, *glnALG*, *fimB*, *fanABC*, *serA* and *ompF* sites, discussed in Fraenkel *et al*. (16), were not used because no experimental data supported binding there.

The **delila**, **alist**, **encode**, **rseq**, **dalvec** and **makelogo** programs were used as described previously to create both natural (Fig. 2) and SELEX (Fig. 5) sequence logos (15,17). The **malign** program was used to adjust and check the alignment of the Lrp sites, and to maximize the information content (18). The **rsim** program was used to determine the standard deviation of $R_{sequence}$ (13,19).

The information content of individual genetic sequences (R_i) can be determined to identify potential binding sites (20–25).The programs **ri**, **scan**, **search**, **live** and **lister** were used to identify and map the Lrp sites relative to the footprint data, and sites were displayed by the sequence walker method (20,21). First, we used **ri** to create an R_{iw} (b, l) weight matrix from the aligned set of sites. Then we scanned each sequence with the natural Lrp weight matrix using **scan**. Next, we used the **search** program to identify and mark the footprinted regions on the map. The **live** program was used to create a spectrum color strip to indicate protein binding site orientation on B-form DNA (Figs 3 and 4). We then

*To whom correspondence should be addressed. Tel: +1 301 846 5581; Fax: +1 301 846 5598; Email: toms@ncifcrf.gov

+Present address: Easton Hall, University of Maryland, College Park, MD 20742, USA

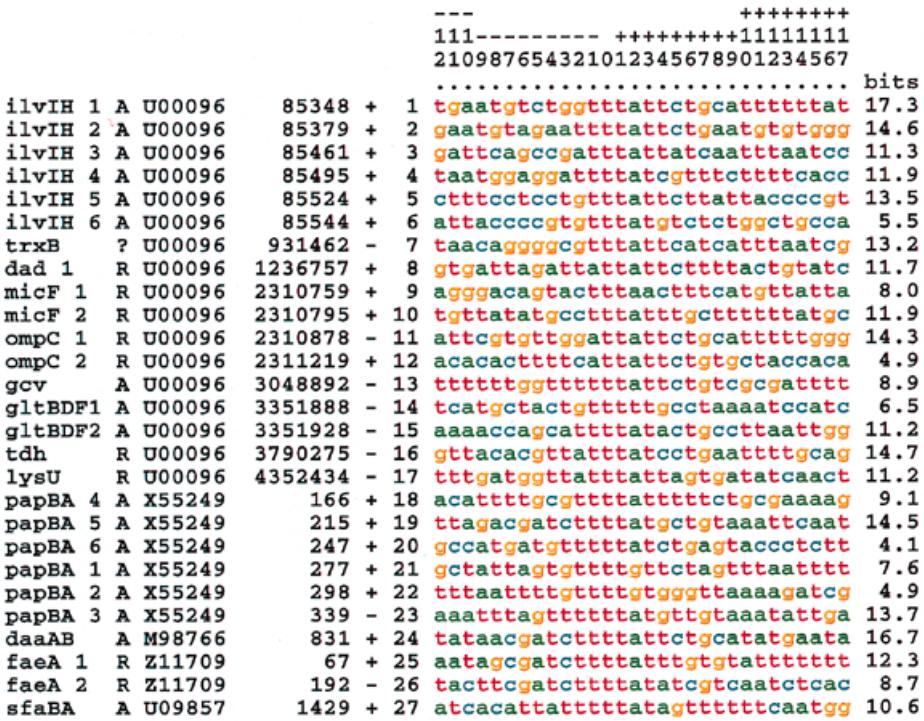

Figure 1. Aligned listing of 27 *E.coli* Lrp binding sites. Columns from left to right indicate gene or operon name; Lrp activation (A), repression (R) or whether its effect is unknown (?); GenBank accession number; zero coordinate in GenBank entry; the orientation of the sequence relative to the GenBank entry; sequence number; the binding sequence and R_i of each site in bits for the range –1 to +12. Twenty-five footprinted sites [*ilvIH* (5,27), *trxB* (36), *micF* and *ompC* (37), *gcv* (38), *gltBDF* (33), *lysU* (39), *papBA* (4,40,41), *faeA* (42), *daaAB* and *sfaBA* (43)], and two mutated sites that affected binding [*dad* (3) and *tdh* (44)] are shown. The alignment is derived from Fraenkel (16). These sites are summarized as sequence logos in Figure 2.

used the **lister** program to place the walkers and other features on the DNA map.

To compare binding energy with individual information (Fig. 6), we plotted the relative binding strength from Cui *et al.* (7) against the strongest R_i found by scanning the SELEX sites with a natural weight matrix. The reported binding site strength was also compared with the R_i predicted from the SELEX sequences themselves. The **xyplo** program was used to generate the graph.

Further information on programs is available at http://www-lecb.ncifcrf.gov/~toms/

RESULTS

Natural Lrp sites

The Lrp sequence logo (Fig. 2A) shows well conserved bases at positions –1 to +3 and low conservation up to position +12. Positions –1 to +3 have heights >1 bit, suggesting major groove binding in that region or minor groove binding with distortion of the helix (26). The cosine wave represents B-form DNA. The sequence logo follows the wave fairly well, a common attribute of many logos in which the peaks correspond to major groove binding (17). However, methylation protection is observed outside of positions –1 to +3, which is inconsistent with this suggested major groove binding. DMS modifies the N7 position of guanine, but these are rare in the region –1 to +3 and none were in the experimental DNA (27), so the DMS protection experiments did not address the question of whether contacts are made in this region. It is possible that Lrp binds to the major groove in the protected regions –5, –4 and +4 to +7 and that the sequence conservation in –1 to +3 is through the minor groove. To account for the sequence conservation exceeding 1 bit in the minor groove, the DNA would have to be heavily distorted, as with TATA binding sites (28–30). The sites contained a total information content of $R_{sequence} = 10.8 \pm 0.9$ bits, for the range –1 to +12.

Lrp is known to both activate and repress transcription (6), so sequence logos for both Lrp activation and repression sites were made (Fig. 2B and C). There are no major differences observed between the sequence characteristics of activation and repression sites, except for more strongly conserved bases at the –10, –9, –2, –1 and +6 positions in the activation logo and a more strongly conserved A at the +2 position and T at the –4 and +12 positions in the repression logo. Activation sites occur about twice as frequently as repression sites.

To see if activation sites can be distinguished from repression sites, activation and repression R_{iw} (*b*, *l*) weight matrices were assembled for individual informational analysis of all footprinted Lrp sites. Activation sites were given higher R_i evaluation by the activation model and the repression sites were favored by the repression model. To test the predictive capability of these matrices, we repeated this analysis but excluded each site from its own matrix. We found that we could not predict repression versus activation. The failure of this bootstrap test, for all sites, suggests that either the activation and repression sites are essentially identical or that more examples are needed to distinguish between them.

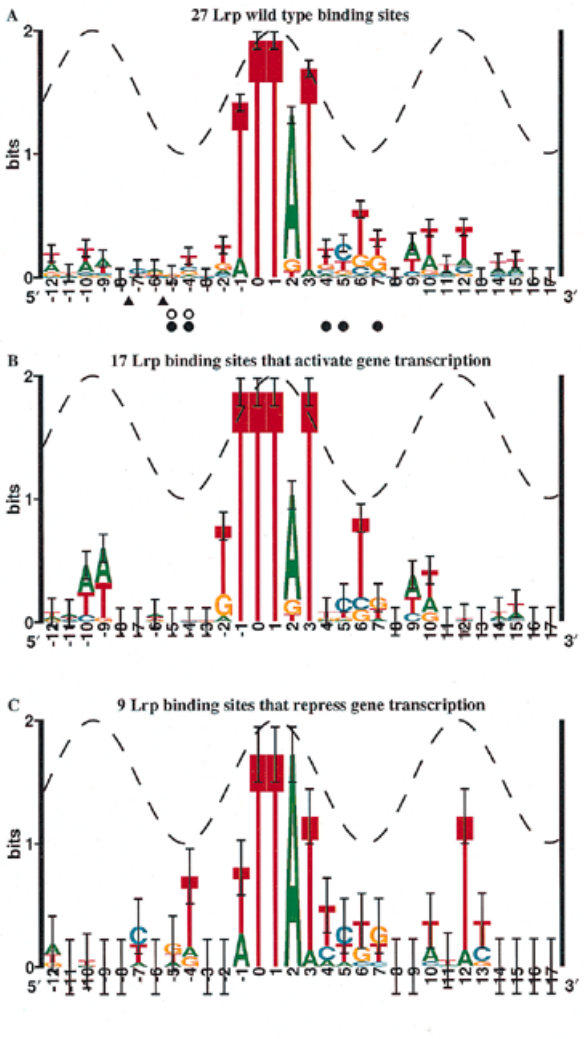

Figure 2. Sequence logos of natural Lrp sites. (**A**) Sequence logo of the 27 wild-type Lrp binding sites shown in Figure 1. Sequence conservation, measured in bits of information, is depicted by the height of a stack of letters for each position in the binding sites. The relative heights of the letters within a stack are proportional to their frequencies. Circles were placed below guanines protected from DMS attacks by Lrp (27). Open circles (○) are guanines protected on the top strand, and filled circles (●) are guanines protected on the bottom strand. Triangles (▲) denote DNase I hypersensitive sites (43). (**B**) Sequence logo of 17 Lrp-activation binding sites (A in Fig. 1). (**C**) sequence logo of nine Lrp-repression binding sites (R in Fig. 1). The cosine wave represents the 10.6 base twist of B-form DNA (17,26).

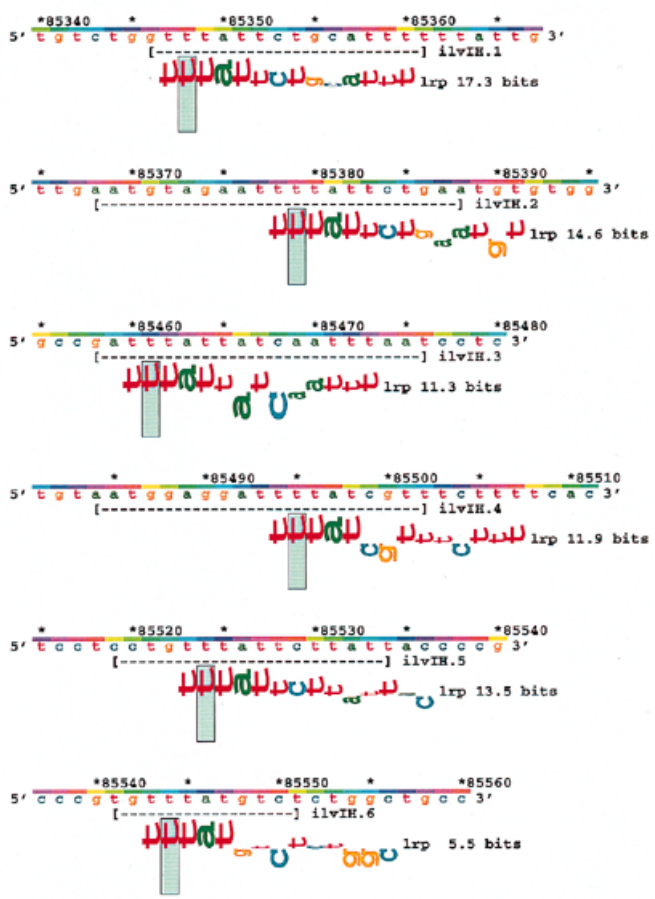

Figure 3. Lrp sequence walkers compared to *in vivo* DMS footprinting data. Six *in vivo* DMS footprinted *ilvIH* sites are marked by dashed lines (5). Beneath these sites are sequence walkers along with the R_i of each site given in bits. The height of each letter in a walker is the sequence conservation that that base contributes to the average sequence conservation shown in the sequence logo (Fig. 2A). The green rectangles mark the zero coordinate of each walker and provide a scale from –3 to +2 bits. All letters of the walkers are rotated 90° counter-clockwise, indicating that all Lrp *ilvIH* sites have the same orientation. Walker location was determined by the **scan** program for R_i > 4 bits, which includes all known natural sites in Figure 1. The asterisks and numbers above the sequence indicate the position on the *E.coli* genome, GenBank entry U00096 (45). The color strip above the sequence has a 10.6 base cycle, representing the helical structure of B-form DNA. Sites 1, 2 and 4 are on the same face because their zero coordinates fall under the same color. Sites 3, 5 and 6 are on the opposite face.

Evidence of model accuracy

To test our model's accuracy, we scanned the complete 27 site individual information weight matrix across the six *in vivo* footprinted *ilvIH* sites (5) and displayed the results using sequence walkers (Fig. 3). The walkers coincided with the six protected areas, confirming the model. The Lrp R_{iw} (*b*, *l*) weight matrix was scanned across the other sites of Figure 1 and similar results were seen (data not shown). Interestingly, the sites all have the same asymmetric orientation and fall on two opposite faces of the DNA.

Informational predictions of possible sites

To test our Lrp binding site model, we excluded the *fimA* regulatory region from our data set (Fig. 1). $(OP)_2Cu^{2+}$ footprinting data along with two-stage methidiumpropyl-EDTA footprinting analysis indicates that Lrp binds to over 60 bp in *fimA* (31,32), and based on a pre-existing consensus sequence, two Lrp binding sites had been predicted at the ends of the protected region (31). As shown by the map of Figure 4, both sites appear to be correctly placed since there is a 10.1 bit walker beneath the first and a 3.3 bit walker partially beneath the second. In addition, the map suggests multiple Lrp binding sites (22) in the *fimA* regulatory region, all in the same asymmetric orientation and on approximately two opposite faces of the DNA.

SELEX-generated sites

Cui *et al*. used the SELEX procedure (8) to obtain sequences that bind Lrp (7). Two SELEX experiments that had been performed

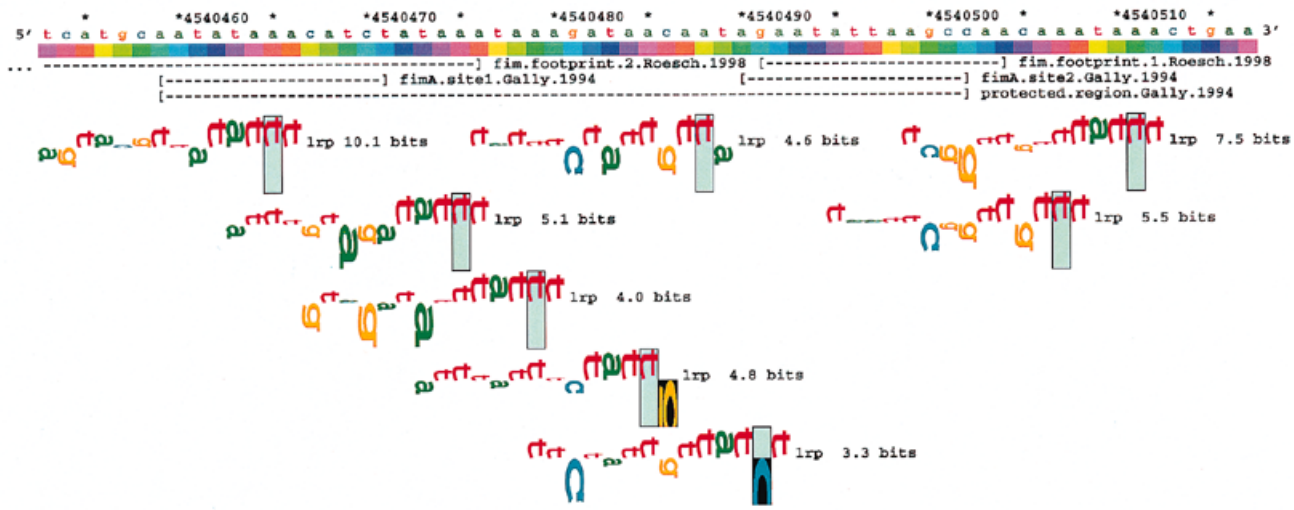

Figure 4. Predicted Lrp sites in the *fimA* regulatory region. Two Lrp binding sites previously predicted (31) are marked with dashes, while $(OP)_2Cu^{2+}$ and two-stage methidiumpropyl-EDTA footprinting data shows extended protection (31,32), that corresponds to a series of Lrp walkers. To include the 3.3 bit site predicted by Gally *et al.* (31), all sequence walkers with $R_i > 3$ bits are shown. The black rectangle indicates a base not observed in the original data set (Fig. 1). If this site were included in the model, it would become 7.2 bits (20). All predicted Lrp sites have the same orientation and the four sites having 10.1, 5.1, 4.8 and 5.5 bits would be on the same face of the DNA since their zero positions are at nearly the same color on the spectrum, while the 4.0, 4.6 and 7.5 bit sites are approximately on the opposite face. The sequence is from GenBank accession U00096 (45).

in the absence of leucine are represented by sequence logos in Figure 5A and B and combined in Figure 5C. A third experiment that had been performed in the presence of leucine has a similar logo (Fig. 5D). Cui *et al*. noticed that the sequences from all three experiments were similar, so they combined all of the sites to generate their consensus sequence. We therefore also made a sequence logo that combined all of the SELEX sequences (Fig. 5E). These sites contained an $R_{sequence}$ of 13.9 ± 0.6 bits, for the range –7 to +8. Various sized symmetric and asymmetric SELEX models were used to predict binding in the natural sequences (Fig. 1) and, unlike the predictions by the natural model, the sites were not consistently identified (data not shown). Finally, we plotted each reported relative binding strength against the corresponding natural R_i values (Fig. 6A), and we plotted the binding strength against R_i values from a weight matrix created from all 67 SELEX sequences (Fig. 6B).

DISCUSSION

Experimentally characterized natural Lrp binding sites (Fig. 1) have typical sequence logos that show moderate variation between activation and repression sites (Fig. 2). However, we were unable to use separated activation and repression models to predict activation and repression. Because they are asymmetric, all natural sequence logos are consistent with a monomer of Lrp binding the DNA, even though experimental evidence suggests homodimer formation in solution (2). While the sinusoidal shape of the logo and sequence conservation in excess of 1 bit suggest major groove binding (17), protection of major groove N7 moieties is outside the region of conservation, suggesting the alternative possibility of minor groove binding in conjunction with large distortions of the DNA helix (26,28,29). In any case, individual information analysis shows that Lrp sites are easily identified in footprinted regions (Figs 3 and 4). As with *Fis* binding sites (22), predicted Lrp binding sites appear to have specific patterns of binding to opposite faces of the DNA helix. In addition, the sites are all oriented in the same direction. The physiological implications of these structures are unknown.

Surprisingly, the sequence logos for natural Lrp binding sites determined by footprints or mutations (Fig. 2A) do not closely resemble the sequence logos obtained by SELEX (Fig. 5E). The SELEX information content is 13.9 ± 0.6 bits whereas the natural is 10.8 ± 0.9 bits, and these differ significantly ($P < 0.005$ by Student's *t*-test for 92 degrees of freedom). Although the central regions from –1 to +3 resemble each other, the SELEX logo has two additional regions, –7 to –5 and +5 to +7, that are complementary to each other (7) and which account for 54% of the SELEX information. A residue of this is visible in the natural logo but only comprises 13% of the natural information. Finally, a lack of correlation between the measured relative binding strength and the R_i found using the natural R_{iw} (b, l) weight matrix (Fig. 6A), demonstrates the incongruity of the two models. Unexpectedly, there was also a poor correlation between the binding strength and an R_{iw} (b, l) weight matrix created from the SELEX sequences themselves (Fig. 6B).

To explain these major discrepancies between the natural and the SELEX sites, we suggest that three proteins are binding in the SELEX experiment, since there are three bulges of sequence conservation that rise above the 1 bit mark in the SELEX logo (Fig. 5E). Two of these are the complementary regions separated by 10 bp at –7 to –5 and +5 to +7, which Cui *et al*. suggested could be bound by a homodimer (7). The correlation between these conserved regions and the 10.6 base cosine wave (Fig. 5E) (17,26), suggests that the homodimer binds in two major grooves on one DNA face. The sequence in between these two conserved regions is asymmetric and vaguely resembles the natural logo (Fig. 2) because of the prominent Ts, suggesting a monomer binding in the major groove of the opposite DNA face.

A possible explanation for differences between the *in vivo* and *in vitro* results is that Lrp naturally forms a trimer with only the central molecule specifically binding to the DNA. A homodimer flanking a central monomer would not only be consistent with

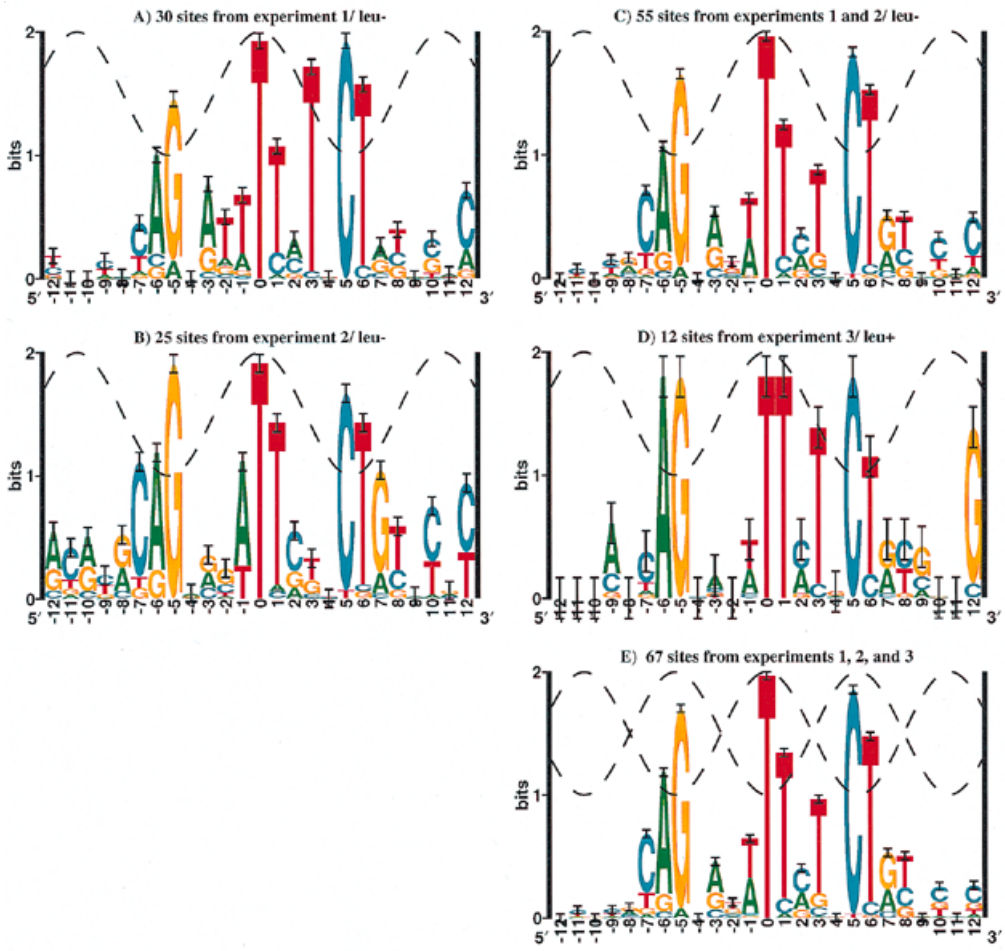

Figure 5. Sequence logos of SELEX-generated Lrp binding sites. (**A**) Sequence logo of 30 sites that made up SELEX experiment 1 of Cui *et al*. (7), in which there was no leucine added. (**B**) Sequence logo of 25 sites that made up experiment 2, without leucine. (**C**) Sequence logo for experiments 1 and 2 combined, without leucine. (**D**) Sequence logo of 12 sites that made up experiment 3, with leucine. (**E**) Sequence logo for experiments 1, 2 and 3 combined. (E) has two cosine waves to show possible major groove binding on two different faces. The SELEX and natural coordinate systems were chosen to facilitate comparison with each other.

sequence conservation in the center of both natural and SELEX logos (–1 to +3 in Figs 2A and 5E) but also explains the methylation protection observed outside the center region in the natural sites and the outer conservation observed in the SELEX sites. Others have suggested that Lrp forms a dimer but that only one of the monomers binds to the DNA (33).

When *in vitro* selections for OxyR and TrpR were analyzed by information theory, they were also found to give results that differ from those obtained from naturally selected sites (17,34). The differences between the *in vivo* and *in vitro* Lrp logos might be attributed to unnatural experimental conditions. Inappropriate salt levels or temperatures, the absence of spermidine (35), or selection of band-shifted DNA with the highest molecular weight (i.e., a triplet complex), among many other possibilities, might be reasons for these results. Sequence logos could be used to quantitatively investigate the effects of such varying conditions (17).

In vitro selection procedures do not always mimic natural evolution (10). The strongest sites, such as those found by SELEX, are not 'optimal' when viewed on an information theory scale (20). Instead, natural sites are observed to have a Gaussian distribution that peters out at the high end. From this viewpoint, the strongest possible sites are seen as abnormal. When SELEX is pushed to obtain the strongest possible binding sites, the resulting sequence logo should show more sequence conservation than the natural sites, and as shown in this paper may be radically different from the natural logo. When the *in vitro* selections are more mild, the logos may resemble each other if the conditions are comparable. If one's goal is to obtain the strongest binder, as has been the emphasis for most of the work with SELEX (10), then strong selection is appropriate and sequence logos can be used to characterize the strong sites. If, instead, the goal is to learn more about the binding pattern of natural ligands, then weaker selection under various conditions could be guided by using sequence logos.

ACKNOWLEDGEMENTS

We thank Paul Shultzaberger for providing R.K.S. with a car, Debby Shultzaberger for re-typing sequences for comparison, Kay Kennedy and Emily Moler for running the Werner H. Kirsten Student Intern Program at NCI, Brenda Deener for scientific guidance and instruction, Mike Miller for bringing attention to grammatical errors, Elaine Bucheimer, Karen Lewis, Paul N. Hengen and Peter K. Rogan for their critique. We thank an

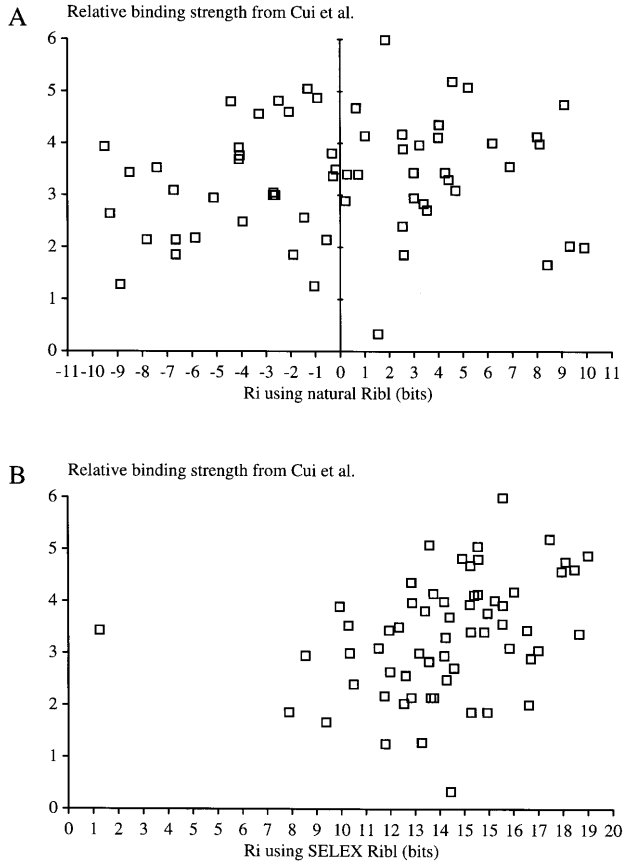

Figure 6. Comparing the relative binding strength of Cui *et al.* with R_i. (**A**) Using the R_{iw} (*b*, *l*) weight matrix from the natural sites (Fig. 1), we scanned the SELEX-generated sequences in Cui *et al*. (7). The highest R_i for each SELEX-generated site was chosen for the 62 sites reported; no correlation was observed between binding strength and R_i (r = 0.15). Also, the sum of all positive R_i values in each sequence was compared with the reported binding energy, but no correlation was found by this second approach (r = 0.06). (**B**) An R_{iw} (*b*, *l*) matrix was made from the SELEX sequences and used to evaluate the same sequences. The single outlier, referred to as 7 in Figure 3 of Cui *et al*. (7), contains a T at +2 that is not observed in any other SELEX sequence and is therefore rated with a low value (20) (r = 0.43 without the outlier).

anonymous referee for suggesting Figure 6B. R.K.S. would also like to thank the National Institutes of Health for funding his research.

REFERENCES

1 Newman,E.B., Lin,R.T. and D'ari,R. (1996) In Neidhardt,F.C. (ed.), *Escherichia coli and Salmonella: Cellular and Molecular Biology.* ASM Press, Washington, DC, pp. 1513–1525.
2 Willins,D.A., Ryan,C.W., Plakto,J.V. and Calvo,J.M. (1991) *J. Biol. Chem*., **266**, 10768–10774.
3 Mathew,E., Zhi,J. and Freundlich,M. (1996) *J. Bacteriol*., **178**, 7234–7240.
4 Braaten,B.A., Plakto,J.V., van der Woude,M.W., Simons,B.H., de Graff,F.K., Calvo,J.M. and Low,D.A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 4250–4254.
5 Wang,Q. and Calvo,J.M. (1993) *J. Mol. Biol*., **229**, 306–318.
6 Calvo,J.M. and Matthews,R.G. (1994) *Microbiol. Rev.*, **58**, 446–490.
7 Cui,Y., Wang,Q., Stormo,G.D. and Calvo,J.M. (1995) *J. Bacteriol*., **177**, 4872–4880.
8 Tuerk,C. and Gold,L. (1990) *Science*, **249**, 505–510.
9 Klug,S.J. and Flamulok,M. (1994) *Mol. Biol. Rep*., **20**, 97–107.
10 Gold,L., Polisky,B., Uhlenbeck,O. and Yarus,M. (1995) *Annu. Rev. Biochem*., **64**, 763–797.
11 Shannon,C.E. (1948) *Bell System Tech. J*., **27**, 379–423.
12 Shannon,C.E. (1948) *Bell System Tech. J*., **27**, 623–656.
13 Schneider,T.D., Stormo,G.D., Gold,L. and Ehrenfeucht,A. (1986) *J. Mol. Biol*., **188**, 415–431.
14 Schneider,T.D. (1994) *Nanotechnology*, **5**, 1–18. http://www-lecb.ncifcrf.gov/~toms/paper/nano2/.Schneider
15 Schneider,T.D. and Stephens,R.M. (1990) *Nucleic Acids Res*., **18**, 6097–6100.
16 Fraenkel,Y.M., Mandel,Y., Friedberg,D. and Margalit,H. (1995) *CABIOS*, **11**, 379–387.
17 Schneider,T.D. (1996) *Methods Enzymol*., **274**, 445–455. http://www-lecb.ncifcrf.gov/~toms/paper/oxyr
18 Schneider,T.D. and Mastronarde,D. (1996) *Discrete Applied Mathematics*, **71**, 259–268. http://www-lecb.ncifcrf.gov/~toms/paper/malign
19 Stephens,R.M. and Schneider,T.D. (1992) *J. Mol. Biol*., **228**, 1124–1136.
20 Schneider,T.D. (1997) *J. Theor. Biol*., **189**, 427–441. http://www-lecb.ncifcrf.gov/~toms/paper/ri/
21 Schneider,T.D. (1997) *Nucleic Acids Res*., **25**, 4408–4415. Erratum (1998) *Nucleic Acids Res*., **26**, 1135. http://www-lecb.ncifcrf.gov/~toms/paper/walker/
22 Hengen,P.N., Bartram,S.L., Stewart,L.E. and Schneider,T.D. (1997) *Nucleic Acids Res*., **25**, 4994–5002. http://www-lecb.ncifcrf.gov/~toms/paper/fisinfo/
23 Rogan,P.K., Faux,B.M. and Schneider,T.D. (1998) *Human Mutat*., **12**, 153–171.
24 Allikmets,R., Wasserman,W.W., Hutchinson,A., Smallwood,P., Nathans,J., Rogan,P.K., Schneider,T.D. and Dean,M. (1998) *Gene*, **215**, 111–122.
25 Kahn,S.G., Levy,H.L., Legerski,R., Quackenbush,E., Reardon,J.T., Emmert,S., Sancar,A., Li,L., Schneider,T.D., Cleaver,J.E. and Kraemer,K.H. (1998) *J. Invest. Dermatol*., **111**, 791–796.
26 Papp,P.P., Chattoraj,D.K. and Schneider,T.D. (1993) *J. Mol. Biol*., **233**, 219–230.
27 Marasco,R., Varcamonti,M., Cara,F.L., Ricca,E., Felice,M.D. and Sacco,M. (1994) *J. Bacteriol*., **176**, 5197–5201.
28 Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) *Nature*, **365**, 512–520.
29 Kim,J.L., Nikolov,D.B. and Burley,S.K. (1993) *Nature*, **365**, 520–527.
30 Penotti,F.E. (1990) *J. Mol. Biol*., **213**, 37–52.
31 Gally,D.L., Rucker,T.J. and Blomfield,I.C. (1994) *J. Bacteriol*., **176**, 5665–5672.
32 Roesch,P.L. and Blomfield,I.C. (1998) *Mol. Microbiol*., **27**, 751–761.
33 Wiese,D.E., Ernsting,B.R., Blumenthal,R.M. and Mathews,R.G. (1997) *J. Mol. Biol*., **270**, 152–168.
34 Haran,T.E. (1998) *J. Biomol. Struct. Dyn*., **15**, 689–701.
35 Pingoud,A. (1985) *Eur. J. Biochem*., **147**, 105–109.
36 Wang,Q., Wu,J., Friedberg,D., Plakto,J. and Calvo,J.M. (1994) *J. Bacteriol*., **176**, 1831–1839.
37 Ferrario,M., Ernsting,B.R., Borst,D.W., Wiese,D.E., Blumenthal,R.M. and Matthews,R.G. (1995) *J. Bacteriol*., **177**, 103–113.
38 Stauffer,L.T. and Stauffer,G.V. (1994) *J. Bacteriol*., **176**, 6159–6164.
39 Lin,R., Ernsting,B., Hirshfield,I.N., Mathews,R.G., Neidhardt,F.C., Clark,R.L. and Newman,E.B. (1992) *J. Bacteriol*., **174**, 2779–2784.
40 Nou,X., Braaten,B., Kaltenbach,L. and Low,D.A. (1995) *EMBO J*., **14**, 5785–5797.
41 van der Woude,M.W., Braaten,B.A. and Low,D.A. (1992) *Mol. Microbiol*., **6**, 2429–2435.
42 Huisman,T.T. and de Graff,F.K. (1995) *Mol. Microbiol*., **16**, 943–953.
43 van der Woude,M.W. and Low,D.A. (1994) *Mol. Microbiol*., **11**, 605–618.
44 Rex,J.H., Aronson,B.D. and Somerville,R.L. (1991) *J. Bacteriol*., **173**, 5944–5953.
45 Blattner,F.R., Plunkett,G.,III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y. (1997) *Science*, **277**, 1453–1474.