Generation of an antibody with enhanced affinity and specificity for its antigen by protein engineering

S. Roberts, J. C. Cheetham & A. R. Rees

Laboratory of Molecular Biophysics, Department of Zoology, Oxford University, The Rex Richards Building, Oxford 0X1 3QU

A detailed description of the interactions between an antibody and its epitope is necessary to allow an understanding of the way in which antibodies bind to antigenic surfaces presented by foreign molecules. Ideally this should be done by analysis of crystal structures of antibody–antigen complexes, but so far only two of these are available[1,2]. An alternative strategy[3] combines molecular modelling[4–6] with site-directed mutagenesis (SDM) and using this we have generated a preliminary model[7] of the complex between Gloop2, an antibody raised against a peptide containing the 'loop' determinant of hen egg-white lysozyme (HEL) which also binds the native protein[8], and its epitope on the protein surface. The main predictions from our model were; (1) that the surface of interaction between the antibody and the antigen is large (20 Å × 15 Å) and involves all the complementarity-determining regions (CDRs), (2) that electrostatic interactions were important in the formation of the complex, and (3) that conformational changes in either the loop or in the CDRs may occur during the formation of the complex. Here we report SDM studies which test some of these predictions; removal of two charged residues at the periphery of the combining site increases the affinity of the antibody for its antigen over 8-fold and decreases its ability to cross-react with closely-related antigens. This result is at variance with our original prediction but can be accommodated within our newly refined model; the role of electrostatics in antigen–antibody interactions is now questionable.

The antibody-combining site is formed by the juxtaposition of six hypervariable or CDRs, three deriving from the light chain (designated L1, L2 and L3) and three from the heavy chain (H1, H2 and H3). Although the structural details of a number of antibody-combining sites are known[9–14], the manner in which CDR-backbone length and sequence influences the final topology is unknown. Such information may eventually be derived from comparisons within a larger data base of structures than presently exists.

Our initial analysis of the computer model of the Gloop2-HEL complex, together with the results of binding studies of Gloop2 and a panel of variant avian lysozymes[8], strongly implicated the interaction of (1) Glu 28 (27A using the Kabat numbering system[15]), in the light chain CDR1 (L1), with Arg 68 (HEL) and (2) Lys 56, in the heavy chain CDR2 (H2), with Asn 77 (HEL). In neither case are the residue pairs close enough to form hydrogen bonds (closest contact 4.7 Å), but it was suggested that they may be important in the orientation of the two interacting protein surfaces. Based on these initial observations both Glu 28 (L1) and Lys 56 (H2) were chosen as candidates for mutagenesis. There were also a number of residues that appeared to be partially buried within the combining site, inaccessible to antigen contact and yet variable between the individual antibodies. Glu 50 in the heavy chain H2 of Gloop2 fell into this category. Four types of substitution were therefore made by SDM: Glu 28 to Ser, Lys 56 to Gln and Glu 50 to Ser and the double mutant Glu 28 to Ser: Lys 56 to Gln according to the scheme shown in Fig. 1, and the mutant proteins were expressed[16].

In parallel with the mutagenesis studies, further refinement of the initial model for the Gloop2–HEL complex was carried out by energy minimization, using the GROMOS molecular dynamics package[17]. Regions of unacceptable stereochemistry in the centre of the combining site region (Fig. 2*a*) were completely eliminated from the model by this procedure, in addition to a general enhancement of the complementarity of the antibody and antigen surfaces (Figure 2*b*). The principal contact residues, however, remain unchanged between the old and new models. The energy minimization of the complex showed that only minor conformational changes (<0.5 Å) would be required within the antibody and the antigen for the two molecules to dock together to form a complex. This is consistent with observations from the experimental study of Poljak and co-workers[1].

The results of binding assays performed with the mutant antibodies, using HEL and Pep1 (a loop-containing peptide) as antigens, are shown in Fig. 3*a* and summarized in Table 1. The single mutant Glu 28 to Ser showed a moderate increase in

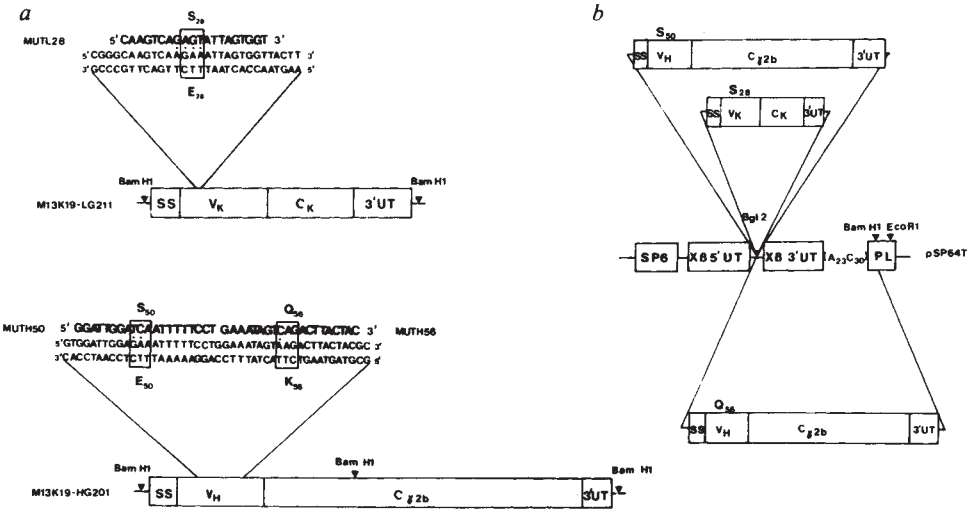

Fig. 1 *a*, The cloning of Gloop2 heavy and light chain complementary DNAs and their expression by *in vitro* transcription with SP6, and *in vivo* translation with *Xenopus* oocytes to generate fully assembled and functional antibodies has been described elsewhere[16]. The same procedure was used to express the mutant antibodies and SDM was performed using the *Eco*K/*Eco*B reciprocating selection system[19]. The cDNA clones were subcloned in an inverted orientation into the M13 mutagenesis vector (M13K19) such that the mutagenic primers were 5′ to the selection primers. Mutagenesis using the mutagenic primer, MUTL28, converted Glu 28(GAA) in CDR1 of the light chain to Ser (AGT) by three mismatches (*), the fourth being silent. The mutagenic primers MUTH50 and MUTH56 generated changes within the CDR2 of the heavy chain at position 50, converting a Glu (GAA) to a Ser (TCA) by two mismatches, and at position 56, a Lys (AAG) to a Gln (CAG) by one mismatch. *b*, Mutant RNA transcripts were synthesized by subcloning the mutated cDNA clones as *Bam*H1 fragments into the *Bgl*II or *Bgl*II/*Bam*H1-restricted vector, pSP64T. All constructs were linearized by *Eco*R1 restriction. The single mutant antibodies, E28S, E50S and K56Q were prepared by comicroinjection of the mutant RNA transcript with the appropriate 'wild-type' heavy or light chain RNA transcript into the cytoplasm of *Xenopus* oocytes; the double mutant E28SK56Q was produced by microinjection of the two mutant transcripts of E28S and K56Q. SS, signal peptide sequence; V, variable region and C, constant

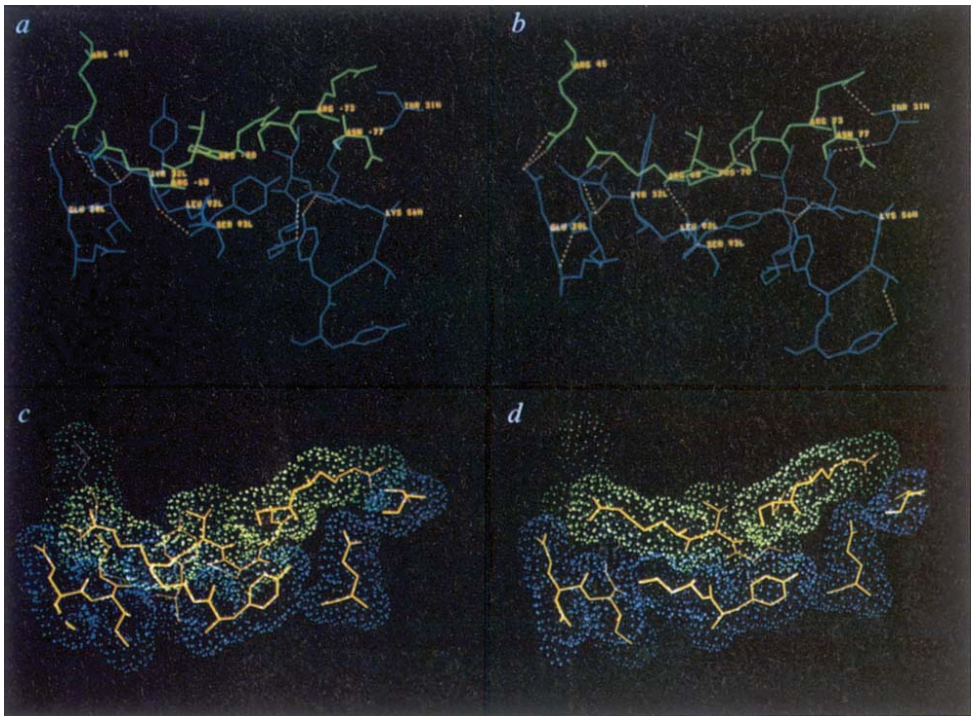

Fig. 2 Combining site region of the Gloop2/HEL complex[7] before (*a*) and after (*b*) refinement of the model by energy minimization. In *a*, note the unacceptably close contacts between (1) main-chain atoms in the region of Leu 92 and Ser 93 of L3, and (2) side-chain atoms in the region of Tyr 32(L1) and Pro 70(HEL). In panel *b*, these bad contact regions are no longer present and the complementarity of the two molecular surfaces is clearly improved. The improvement in the model resulting from the energy minimization was reflected in the large potential energy (p.e.) drop over the 164 cycles of steepest descent minimization: p.e. (initial model) = $+2.95\times10^{11}$ kcal mol^{-1}; p.e. (minimized model) = -0.533×10^{4} kcal mol^{-1}. A van der Waals representation of the antibody (blue): antigen (green) surfaces (*c* shows model *a*, and *d* shows model *b*) indicates the general stereochemical complementarity of the two molecules in the region of the combining site. Hydrogen bonds, both within the individual molecules themselves, and between antigen and antibody, are indicated (- - -).

Table 1 Affinity constants (k_D/M) for binding of Gloop 2 and its mutants to Pep1 and HEL

	Gloop 2	E28S	K56Q	E28SK56Q
Pep1	$2.0\pm0.75\times10^{-8}$	8.2×10^{-9}	2.0×10^{-8}	4.8×10^{-9}
HEL	$2.3\pm1.0\times10^{-7}$	4.7×10^{-8}	$1.9\pm0.12\times10^{-7}$	$2.75\pm0.02\times10^{-8}$

Assays were carried out in triplicate as described in Roberts and Rees[16]. Essentially microtitre plates were coated with goat anti-mouse IgG antibody (affinity purified) at a concentration of 50 μg ml^{-1} in PBS for 18 h at 4 °C. After coating the plates were blocked with PBS containing 0.05% Tween for 30 min at 4 °C, then 2 × 1 min at room temperature. Oocyte test supernatant was added and incubated at room temperature for 6 h. Plates were washed and ^{125}I-labelled Pep1 with and without inhibitor was added and incubated at room temperature for a further 6 h. Plates were washed with PBS/Tween 3 × 1 min and individual wells cut out and counted. The k_D values for Pep1 inhibition of ^{125}I-labelled Pep1 binding were obtained from Scatchard analysis. As Scatchard analysis was inappropriate for analysis of HEL inhibition of Pep1 binding, k_D values were obtained by the following method. Using the program SANCOL (R. Ryan, unpublished) a non-linear regression procedure[22] was used to fit the binding equation which relates specifically bound units (for instance c.p.m.—the measured variable) and the total ligand concentration (the control variable) to the experimental data [23]. Scatchard estimates of the k_D and total site concentration (S_T) values were then used as a start point for an iterative procedure which calculated the final k_D and S_T values. Binding data were measured at *p*H 7.2. The range of means obtained in different experiments is given.

affinity for both Pep1 and HEL (3-4 fold) whereas the mutant Lys 56 to Gln showed no significant change in binding. But combining the two single mutations within the same antibody gave a double mutant which showed a marked increase in affinity for HEL (8-9-fold), and a moderate increase for Pep1 (4-5-fold) (Fig. 3*a*).

When this analysis was extended to the variant lysozymes (Table 2) the pattern of binding satisfied the purely thermodynamic criteria that, if the mutations improved the complementarity of Gloop2 for its native antigen HEL, then the variant lysozyme with the highest affinity, but non-identical complementarity, for wild type Gloop2 should experience the most drastic loss of affinity (6-7-fold decrease in relative affinity for TEL, Table 2). By contrast, those variant lysozymes with somewhat lower affinities, and hence rather less stringent interactions,

Table 2 Relative affinities of Gloop 2 and its mutants for lysozymes of different species

	Gloop 2	E28S	K56Q	E28SK56Q
TEL	0.5	3.0	0.5	7
BWQEL	124	111	ND	235
RNPEL	13	ND	12	30
HuL	370	ND	ND	1300

The numbers shown express ID_{50} values for inhibition of binding of ^{125}I-labelled Pep 1 to antibody by the appropriate variant lysozyme (methods as described in the legend to Table 1). Each value has been normalized to the ID_{50} value for binding of the appropriate antibody to HEL. Variant lysozymes used in this study are: TEL, turkey egg lysozyme; BWQEL, bobwhite quail egg lysozyme; RNPEL, ring-necked pheasant egg lysozyme, and HuL, human lysozyme.

showed not only a greater affinity for HEL, but also an increased specificity towards HEL over the other avian species. Modelling of these amino-acid changes in the antibody L1 and H2 CDRs suggests that the removal of the electrostatic residues, and their replacement by non-charged hydrophilic residues, allows a closer fit of antibody and antigen. As a consequence there is a general improvement in hydrogen bonding between the two surfaces, but without significant changes in the interactions within the antibody CDR's at the sites of the mutations (Fig. 3*b*).

By contrast, mutation of the partially-buried Glu 50 in H2 resulted in abolition of binding between the antibody and HEL or Pep1 (Fig. 3*c*). The residue in this position is the first to emerge from the framework region preceding CDR H2 and shows a high variability between heavy chains of the same V_HII subgroup[15]. In the model this residue appears to be both hydrogen bonded to Tyr 94 in neighbouring CDR L3 and involved in a salt-bridge-like interaction with Arg 101 of H3. Substitution at this position would be predicted to cause a significant change in the relative conformations of adjacent, interacting CDR's. Such perturbations, if they occurred, could more than account for the observed loss of binding. This observation may be an example of a situation where the existence of variability between

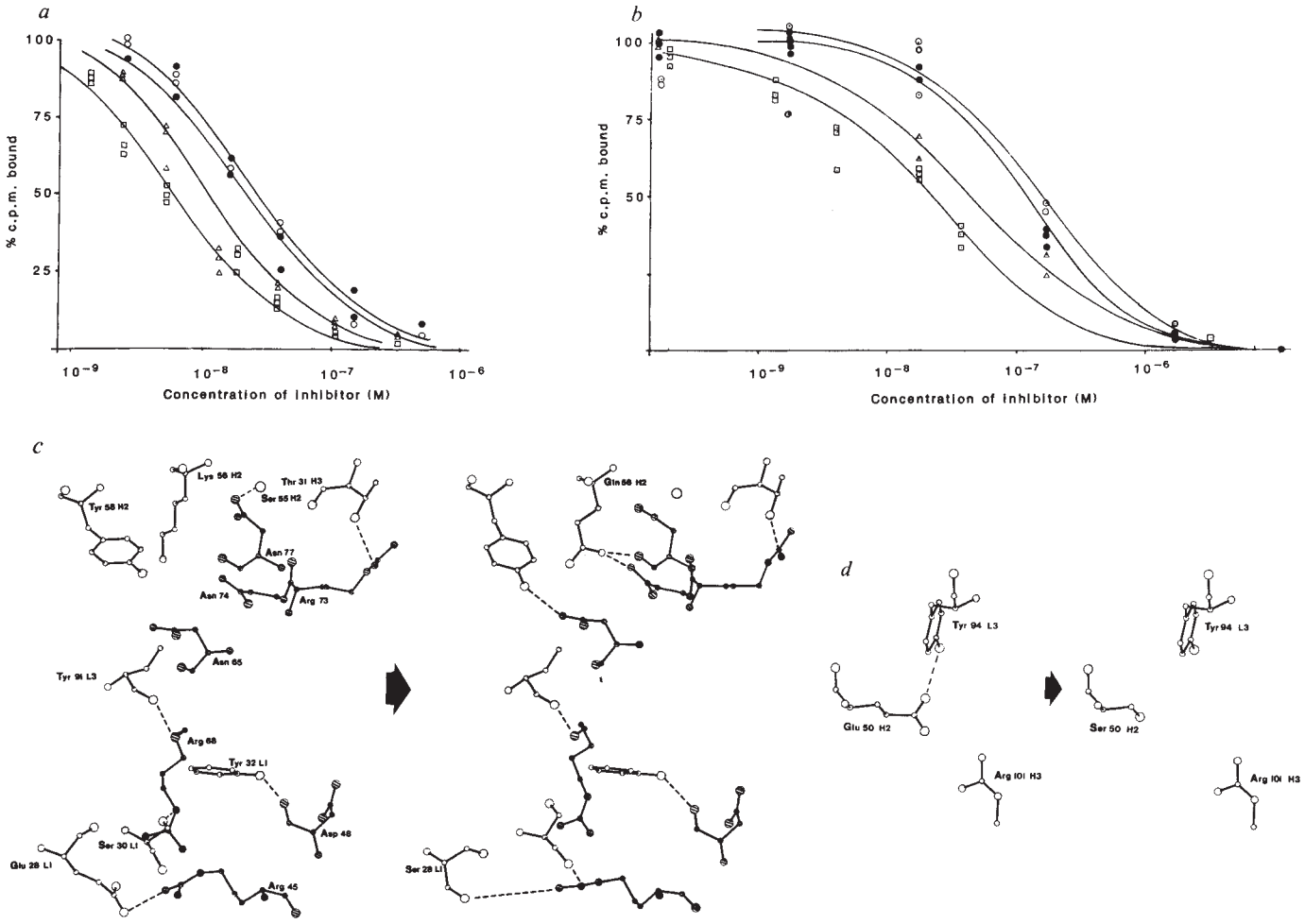

Fig. 3 Inhibition of binding of ^{125}I-labelled Pep1 to Gloop2 (○) and mutant antibodies, k56Q (●), E28S (△) and E28S K56Q (□) by *a*, Pep1 and *b*, HEL. *c*, Predicted effect of the combined mutations Glu 28-Ser (L1) and Lys 56-Gln, in the region of the antibody combining site. The modelling would suggest that hydrogen bonding interactions (- - - -) within the antibody CDR would be preserved, and in addition interactions with the antigen enhanced; Asn 65(HEL) → Tyr 58(H2), Asn 74(HEL) → Gln 56(H2), Asn 77(HEL) → Gln 56(H2), Arg 45(HEL) → Ser 28(L1) and Arg 45(HEL) → Ser 30(L1) all appear as new contacts. *d*, Predicted effect of the mutation Glu 50 → Ser (H2). Here the modelling suggests the replacement of the glutamic-acid residue by a serine will have a significant effect on the nature of the antibody-combining site, with the loss of two important interactions: (1) a hydrogen bond between Tyr 94 OH(L3) and Glu 50 OE1 (H2) and (2) a salt-bridge-like interaction between Glu 50 (H2) and Arg 101(H3). Amino-acid changes were incorporated into the model by introducing the new side chains with FRODO[20] on the PS300, and then energy minimizing the entire structure, using GROMOS[17], to obtain a model for the mutant protein[21].

site. Different combinations of CDR may have requirements for specific inter-CDR interactions. Thus, when assessing the structural consequences of somatic mutations *in vivo* or *in vitro*, both antigen–antibody and CDR–CDR interactions should be considered.

These preliminary results therefore raise two important questions concerning antibody–antigen interactions, and suggest a number of possible mutations for further study. Our original premise was that the two charged groups lying at opposite edges of the combining site (Glu 28 and Lys 56) were important contacts in the antigen interaction and, further, might actually play a role in orientating the loop region of HEL. But the engineering of an antibody with enhanced affinity for its antigen by the removal of these proposed 'key' electrostatic residues, and their replacement by non-charged polar residues, questions this hypothesis. Our results would indicate that the hydrogen-bonding between the two molecules over the combining site surface is of paramount importance. The electrostatic residue in the antigen (Arg 68), no longer paired across the interface after mutagenesis, is easily accessible to solvation by water over the ionic strength range 0.01 M to 0.5 M, such as might be expected if electrostatic orientation was important in acquisition of the complex, (2) the k_D for binding of Pep1 to Gloop2 actually decreases with ionic strength from 1.15×10^{-8} M (at 0.01 M) to 3.8×10^{-9} M (at 0.50 M). As k_{on} is unchanged this result indicates that the charged residues actually exert an inhibitory effect on binding, and that when their effective charge is screened improved protein–protein contacts are possible. This proposition is consistent with the mutation experiment where the removal of the charges results in increased affinity. The possible counter-argument that, by substitution of two charged residues the surface has been rendered more 'sticky' by altering its hydrophobic/hydrophilic character is not tenable because solvent accessibility calculations[18] suggest that the hydrophilic character is largely unchanged by the substitutions Lys 56 to Gln and Glu 28 to Ser (SA < 30 Å^2).

In conclusion, we have demonstrated that it is possible to engineer an anti-peptide antibody in such a way that its affinity for the same epitope in the native protein is increased and, concomitantly, its cross-reactivity with related antigens is

This approach offers a possible solution to the problem of how to generate high affinity antibodies against intact antigens when peptides are used as immunogens and thus has far-reaching therapeutic consequences.

We thank the Science and Engineering Research Council (SERC) (S.R.) and the SERC/Industry Protein Engineering Programme (J.C.) for financial support. We also thank Stuart Bradford for help with the generation and characterization of one of the mutant antibodies, Professor R.J.P. Williams for useful discussions concerning the analysis of the binding data and Professor Sir David Phillips for continued support.

Received 26 March; accepted 2 June 1987.

1. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, P. J. *Science* **233,** 747-753 (1986).
2. Colman, P. M. *et al. Nature* **326,** 358-363 (1987).
3. Rees, A. R. & de la Paz, P. *Trends Biochem.* **123,** 144-148 (1986).
4. Davies, D. R. & Padlan, E. A. in *Antibodies in Human Diagnosis and Therapy* 119-132 (Raven, New York, 1976).
5. Greer, J. *J. molec. Biol.* **153,** 1027-1042 (1981).
6. Novotny, J., Bruccoleri, R. & Karplus, M. *J. molec. Biol.* **177,** 787-818 (1984).
7. de la Paz, P., Sutton, B. J., Darsley, M. J. & Rees, A. R. *EMBO J.* **5,** 415-425 (1986).
8. Darsley, M. J. & Rees, A. R. *EMBO J.* **2,** 383-392 (1985).
9. Segal, D., Padlan, E., Cohen, G., Rudikoff, S., Potter, M. & Davies, D. *Proc. natn. Acad. Sci. U.S.A.* **71,** 4298-4302 (1974).
10. Epp, O., Lattman, E., Schiffer, M., Huber, R. & Palm, W. *Biochemistry* **14,** 4943-4952 (1975).
11. Saul, F., Amzel, L. & Poljak, R. *J. biol. Chem.* **253,** 585-597 (1978).
12. Marquart, M., Deisenhofer, J., Huber, R. & Palm, W. *J. molec. Biol.* **141,** 369-391 (1980).
13. Furey, W., Wang, B. C., Yoo, C. S. & Sax, M. *J. molec. Biol.* **167,** 661-692 (1983).
14. se Won Suh, *et al. Proteins* **1,** 74-80 (1986).
15. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. *Sequences of Proteins of Immunological Interest.* (U.S. Department of Health and Human Services, National Institutes of Health, Bethesda, 1983).
16. Roberts, S. R. & Rees, A. R. *Protein Engng* **1,** 59-65 (1986).
17. Aqvist, J., van Gunsteren, W. F., Leijonmarck, M. & Topia, O. *J. molec. Biol.* **183,** 461-477 (1985).
18. Richards, F. M. *Adv. Biophys. Bioeng.* **6,** 151-176 (1977).
19. Carter, P., Bedouelle, M. & Winter, G. *Nucleic Acid Res.*, **13,** 4431-4443 (1985).
20. Jones, T. A. *Computational Crystallography* 303 (Clarendon, Oxford, 1982).
21. Shih, H. H. L., Brady, J. & Karplus, M. *Proc. natn. Acad. Sci. U.S.A.* **82,** 1697-1700 (1985).
22. Duggleby, R. G. *Analyt. Biochem.* **110,** 9-18 (1981).
23. Munson, P. J. & Rodbard, D. *Analyt. Biochem.* **107,** 220-239 (1980).

Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region

K. Diane Jofuku, Jack K. Okamuro & Robert B. Goldberg*

Department of Biology, University of California, Los Angeles, California 90024, USA

Seed protein genes are highly regulated during the soybean life cycle[1,2]. These genes encode prevalent mRNAs that accumulate and decay during embryogenesis, and are either undetectable or present at low levels in mature plant organ systems[2,3]. Transcriptional activation and repression processes are important in regulating seed protein gene developmental expression programs[1,2]. We started DNA binding protein studies with the soybean lectin gene[4-6] to begin to identify *trans*-acting proteins and *cis*-regulatory sequences required for seed protein gene expression. We have identified an embryo DNA binding protein that interacts with specific sequences in the lectin gene 5′ region. The DNA binding protein is undetectable in mature plant organ systems and its concentration parallels the lectin gene transcription rate during embryogenesis. The DNA binding protein activity corresponds to a 60,000 M_r (60K) nuclear protein, and a protein of similar size interacts with at least one other seed protein gene but not with a gene inactive during embryogenesis. Our data suggest that the 60K protein, and the DNA sequences that it interacts with, may be involved in regulating lectin gene expression.

Figure 1*a* schematically shows the lectin gene region. We isolated this region as a 17.1-kilobase (kb) *Eco*RI fragment from a λ Charon 4 soybean genomic library[4,7]. In addition to the lectin gene, the 17.1-kb fragment contains at least four nonseed protein genes[6]. Figure 1*b* shows the lectin gene and relevant reference sequences[5]. The lectin gene is intronless, encodes a 1.1-kb mRNA, and is expressed during specific embryonic periods and in the mature plant root[6]. *In situ* hybridization studies showed that lectin mRNA is represented primarily in embryo cotyledon cells and in root ground meristem tissue (L. Perez-Grau and R.B.G., unpublished data). By contrast, the nonseed protein genes are expressed throughout embryogenesis and in mature plant leaf, root, and stem cells[6]. Thus the lectin gene is regulated temporally and spatially during the soybean life cycle, and is embedded in a domain with several differentially expressed genes.

We isolated nuclear proteins from mid-maturation stage embryos, 75 days after flowering (DAF)[8] and then reacted these proteins with lectin gene fragments to identify *cis*-elements and *trans*-factors that are involved in regulating lectin gene expression. At mid-maturation, cell division has ceased, embryo cells are expanding and accumulating seed proteins, and lectin mRNA is ~0.75% of the embryo mRNA mass[4,8]. Figure 1*c* shows the results of a DNA gel electrophoresis mobility retardation assay[9] using a lectin gene 5′ probe (Lel 5′, Fig. 1*b*). We used this probe initially because gene transfer studies showed that the lectin gene is regulated correctly in tobacco[6], and that 0.50 kb of 5′ sequence (pLe1Δ0.5, Fig. 1*a*) is sufficient to program its expression during seed development (J.K.O. and R.B.G., unpublished results). As seen in Fig. 1*c*, Lel 5′ probe mobility was retarded significantly in the presence of nuclear proteins indicating that protein–DNA complexes formed (lanes NP and O). Addition of unlabelled poly(dI-dC)·poly(dI-dC) duplex DNA to eliminate nonspecific protein–DNA interactions[9] yielded a free Lel 5′ fragment (circle, Fig. 1*c*) and a more slowly migrating protein–DNA complex (arrow, Fig. 1*c*). Figure 1*d* shows that we were unable to detect a protein–DNA complex with a lectin gene 3′ probe (Lel 3′, Fig. 1*b*). Nor was a complex detected with a gene that is not expressed in soybean embryos (leghaemoglobin 5′, Fig. 1*d*). We added unlabelled lectin DNA (pH4.4, Fig. 1*a*), as well as unlabelled pBR322 and *Drosophila* blastoderm gene DNAs, to determine if the protein–DNA complex was specific for the lectin gene. Only the unlabelled lectin DNA eliminated the protein–DNA complex formation (Fig. 1*e*). Together, these data show that embryo nuclear protein forms a specific complex with the lectin gene, and that the protein–DNA interaction occurs in a 5′ gene region.

We reacted the Lel 5′ probe (Fig. 1*b*) with nuclear proteins from embryos at different developmental stages, and from mature plant organ systems, to test whether the DNA binding protein activity correlated with lectin gene transcription levels. The mobility retardation assay in Fig. 2*a* shows that a protein–DNA complex (arrow) was obtained with the nuclear protein extract of 25-DAF embryos (lane DAF 25) demonstrating that lectin DNA binding protein activity is present early in embryogenesis. We also obtained a protein–DNA complex with similar electrophoretic mobility using the nuclear protein extract from 40-DAF embryos (Fig. 2*a*, lane DAF 40); however, the proportion of Lel 5′ probe in the complex increased significantly. By contrast, the fraction of Lel 5′ probe in the protein–DNA complex decreased with the 75-DAF embryo nuclear protein extract (Fig. 2*a*, lane DAF 75) and was reduced to an undetectable level with nuclear proteins from embryos in the terminal stages of development (Fig. 2*a*, lane DAF 100). In addition, we were unable to detect a protein–DNA complex with leaf, stem, and root nuclear proteins (Fig. 2*a*, lanes L, S, R).

We presented elsewhere the relative lectin gene transcription rates and mRNA levels during different stages of the soybean life cycle[2,6]. Lectin mRNA accumulates and decays during embryogenesis and is undetectable in mature plant leaf and