

## A Zinc-binding Domain Involved in the Dimerization of RAG1

Karla K. Rodgers<sup>1</sup>, Zimei Bu<sup>1</sup>, Karen G. Fleming<sup>1</sup>, David G. Schatz<sup>2</sup>  
Donald M. Engelman<sup>1</sup> and Joseph E. Coleman<sup>1\*</sup>

<sup>1</sup>Department of Molecular Biophysics and Biochemistry  
Yale University, 260 Whitney Avenue, P.O. Box 208114  
New Haven, CT 06520-8114  
USA

<sup>2</sup>Howard Hughes Medical Institute, Section of Immunobiology, Yale University School of Medicine, 310 Cedar Street  
P.O. Box 208011  
New Haven, CT 06520-8011  
USA

Recombination-activating gene 1 (*RAG1*), as well as *RAG2*, are the only lymphoid-specific genes required for V(D)J recombination. *RAG1* protein contains a C<sub>3</sub>HC<sub>4</sub> zinc-binding motif (zinc ring finger) that binds two zinc ions. We have found that *RAG1* contains additional zinc-binding motifs in the form of two separate C<sub>2</sub>H<sub>2</sub> zinc finger sequences. One of the zinc fingers, in combination with the C<sub>3</sub>HC<sub>4</sub> subdomain, forms a highly specific dimerization domain. A combination of biophysical techniques has been used to determine the energetics of association, the overall shape of the dimerization domain, and the relative orientation of the monomeric subunits within the dimer. These results provide direct evidence that a C<sub>3</sub>HC<sub>4</sub> motif is involved in a protein–protein interaction, in this case *via* homodimer formation. In addition, the observation that the dimerization domain includes multi-class zinc binding motifs, namely both a zinc finger and a C<sub>3</sub>HC<sub>4</sub> subdomain, has important implications for other C<sub>3</sub>HC<sub>4</sub>-containing proteins. The position of this dimerization domain in the N-terminal third of the *RAG1* sequence of 1040 amino acid residues may have a significant influence on the activities associated with the C-terminal domains of the protein.

© 1996 Academic Press Limited

**Keywords:** *RAG1*; V(D)J recombination; C<sub>3</sub>HC<sub>4</sub> zinc domains; small-angle X-ray scattering; analytical ultracentrifugation

\*Corresponding author

### Introduction

The tremendous diversity of the variable regions of T cell receptor and immunoglobulin chains originates during lymphocyte development through a complex process referred to as V(D)J recombination. This site-specific recombination event selectively combines in an ordered process gene segments known as V (variable), J (joining), and sometimes D (diversity) to produce the genetic sequence encoding the variable region of the antigen binding proteins (reviewed by Lewis, 1994). The site specificity arises as each gene segment is flanked by a recombination signal sequence (RSS) containing conserved heptamer and nonamer DNA sequences. Two types of RSSs are required for an efficient V(D)J recombination reaction, one with

heptamer and nonamer separated by 12 bp, the other with heptamer and nonamer separated by 23 bp. Site-specific cleavage occurs between the coding region of the respective V, D, or J gene segment and the heptamer of the RSS. Products of the reaction are a coding joint, consisting of imprecisely joined gene segments, and a precise heptamer-to-heptamer signal joint. The reaction results in either deletion or inversion of the intervening DNA separating the coding segments. The nature of these signal and coding joints is well-characterized.

It is now believed that V(D)J recombination occurs through the combined action of multiple proteins, including both ubiquitously expressed factors and those specific to lymphoid cells. The two recombination activating genes, *RAG1* and *RAG2*, are the only essential lymphoid-specific components of the enzymatic machinery (Schatz *et al.*, 1989; Oettinger *et al.*, 1990). Extensive mutagenesis of the RAG proteins has allowed the identification of the essential “core” regions of the proteins required for recombination of

Abbreviations used: *RAG1*, recombination-activating gene 1; *RAG2*, recombination-activating gene 2; RSS, recombination signal sequence; MBP, maltose-binding protein; ZFA, zinc finger A; ZFB, zinc finger B; PCR, polymerase chain reaction.

chromosomal substrates (Sadofsky *et al.*, 1993, 1994; Silver *et al.*, 1993; Cuomo & Oettinger, 1994). More recently, it has been shown that the core regions of RAG1 and RAG2 proteins are necessary and sufficient for recognition and cleavage at a single RSS, demonstrating that the RAG proteins play a central role in the earliest stages of V(D)J recombination (van Gent *et al.*, 1995; McBlane *et al.*, 1995). The additional observation that RAG1 and RAG2 co-localize *in vivo* in a multi-protein complex suggests that these proteins form a distinctive enzymatic complex (Leu & Schatz, 1995; Spanopoulou *et al.*, 1995). Aside from the apparent enzymatic activities of RAG1 and RAG2 in V(D)J recombination, little is known about specific functional or structural aspects of these two proteins. No regions involved in protein-protein interactions, DNA recognition, or even definitive enzymatic sites within either protein have been identified.

One region that may play a significant role in the activity of the full-length protein is a sequence that has been identified as a zinc ring finger subdomain or a C<sub>3</sub>HC<sub>4</sub> zinc binding motif, found within the N-terminal third of RAG1. This motif has been found in over 40 proteins from mammals, plants, fungi, insects, and viruses, including several proto-oncogenes and the human breast cancer susceptibility gene, BRCA1 (Freemont *et al.*, 1991; Freemont, 1993; Futreal *et al.*, 1994). NMR structures have been solved of C<sub>3</sub>HC<sub>4</sub> domains contained in the equine herpes virus gene 63 protein (EHV1-gene 63) and the mammalian putative transcription factor PML (Barlow *et al.*, 1994; Borden *et al.*, 1995). In both structures the metal ligands are interleaved between the two zinc sites *via* a three-stranded antiparallel  $\beta$ -sheet. However, despite basic similarities the two structures differ significantly in detail (Borden *et al.*, 1995). This is in contrast to other classes of zinc binding domains, such as the zinc finger, where structural variations among members of the class are more limited. In spite of its widespread distribution, in no case has the function of a C<sub>3</sub>HC<sub>4</sub> motif been determined.

We have characterized several domains of RAG1 including the C<sub>3</sub>HC<sub>4</sub> motif, immediately contiguous subdomains, and one distant subdomain. In addition to the C<sub>3</sub>HC<sub>4</sub> subdomain we have found two other zinc binding motifs in RAG1, both resembling zinc finger sequences. One of the zinc finger motifs in combination with the C<sub>3</sub>HC<sub>4</sub> subdomain forms a stable dimerization domain as determined by several biophysical techniques. This would be the first example in which two separate classes of zinc binding domains acting in conjunction fulfil a structural and functional role. These results demonstrate that C<sub>3</sub>HC<sub>4</sub> zinc-binding motifs can participate in protein-protein interactions. The position of this dimerization domain immediately adjacent to the core region of the RAG1 protein is consistent with it serving an important function in recombination of antigen receptor genes *in vivo*.

## Results

### RAG1 contains four zinc binding sites

Sequence analysis of RAG1 previously identified a C<sub>3</sub>HC<sub>4</sub> zinc binding motif in the N-terminal third of the protein. We identified by sequence analysis two additional potential zinc binding motifs in the protein, both consisting of C<sub>2</sub>H<sub>2</sub> zinc finger sequences. The positions of the two zinc fingers relative to the C<sub>3</sub>HC<sub>4</sub> motif in RAG1 and their amino acid sequences are shown in Figure 1. The first zinc finger, referred to as ZFA, is adjacent to the C<sub>3</sub>HC<sub>4</sub> motif within the primary sequence and contains the potential zinc ligands Cys353, Cys359 His369 (or His370), and His374. A second zinc finger sequence, referred to as ZFB, is present within the carboxy-terminal third of RAG1 and includes the potential zinc ligands Cys727, Cys730, His744, and His750.

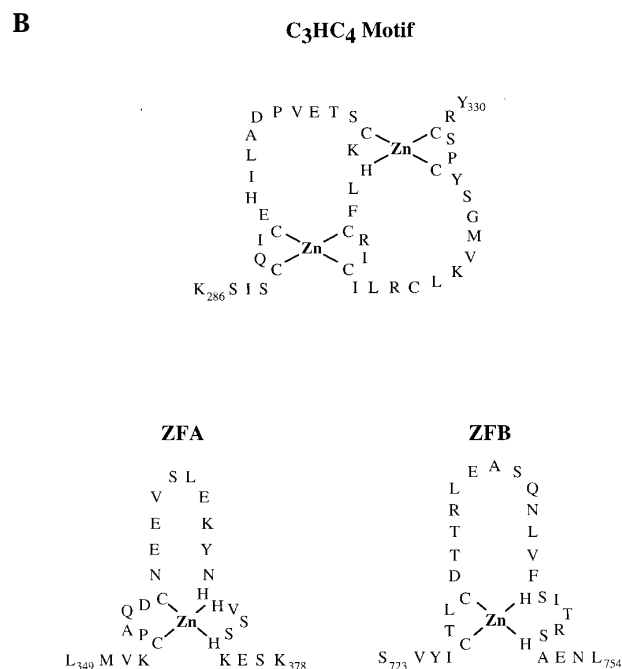
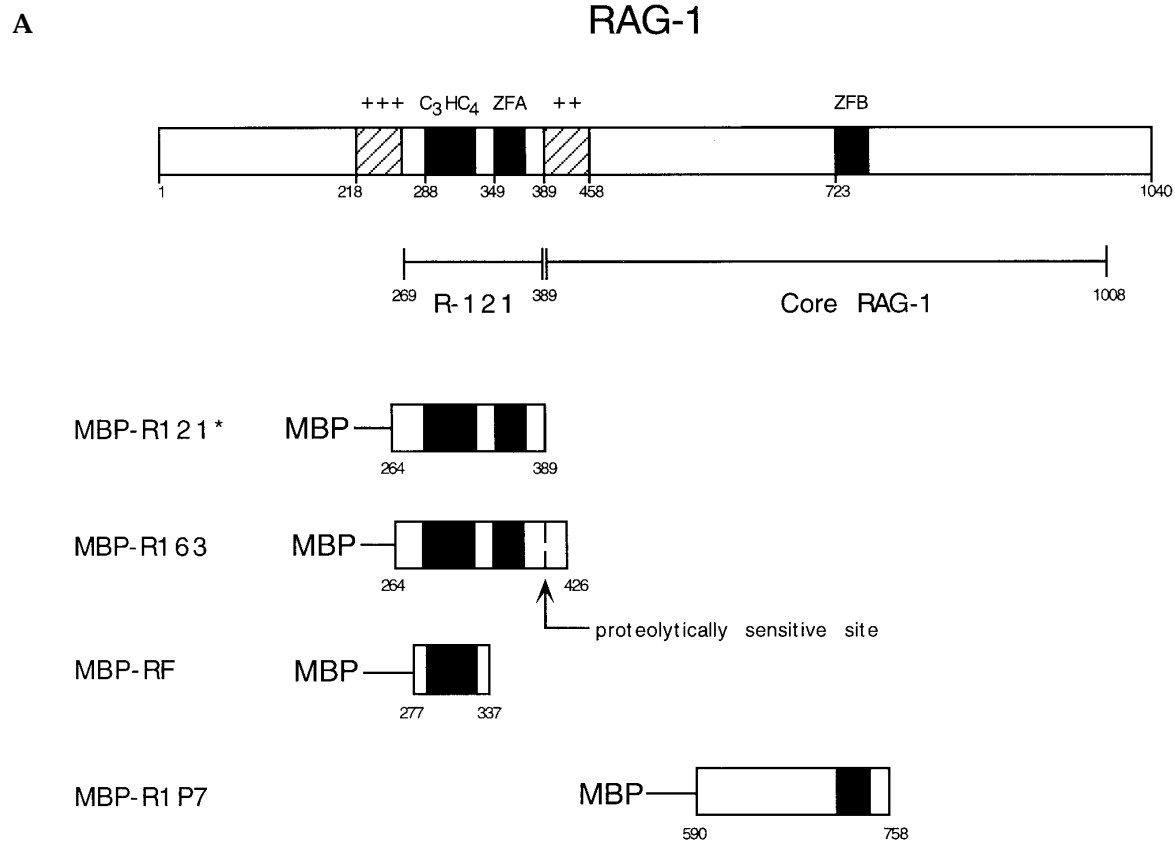
In order to determine the metal to protein stoichiometry present in the subdomains described above, several fragments of RAG1 were cloned and overexpressed in *Escherichia coli* as chimeric proteins with the maltose-binding protein (MBP; Figure 1). MBP was found to enhance expression as well as the solubility of the RAG1 fragments in some cases. The MBP fusion is constructed such that the four C-terminal residues of MBP are replaced by a 20 amino acid residue sequence resulting in a 16 amino acid residue linker between MBP and the cloned sequence (Maina *et al.*, 1988). We used atomic absorption spectroscopy to measure the number of zinc ions bound per RAG1 fragment (Table 1). The MBP-RF construct consists of only the C<sub>3</sub>HC<sub>4</sub> subdomain of RAG1 (Figure 1). As expected this protein binds two zinc ions. The MBP-R163, MBP-R121\* and R121 proteins all bind three zinc ions (Figure 1 and Table 1). The R121 protein was obtained from cleavage of the MBP-R121\* with limited proteolysis by trypsin as described below. All three of these RAG1 fragments contain both the C<sub>3</sub>HC<sub>4</sub> and ZFA subdomains; thus the observation of three bound zinc ions is consistent with two metals coordinated to the C<sub>3</sub>HC<sub>4</sub> motif and the third metal bound to the ZFA subdomain.

To determine if zinc binds to the ZFB sequence we measured the metal stoichiometry of MBP-R1P7 (Figure 1). This protein binds one zinc ion as predicted by the presence of a single zinc finger-like sequence (Table 1). Additional evidence of thiolate ligation within ZFB is provided by the exchange of zinc with cadmium in MBP-R1P7. This results in the appearance of a broad band in the UV absorbance at 245 nm, indicative of a S<sup>-</sup> → Cd charge-transfer band (data not shown). As expected a similar result is seen with MBP-R121\* and R121.

Since zinc binding domains are typically associated with DNA binding, it is possible that one or more of these domains of RAG1 bind to specific DNA sequences. The most likely candidates for

specific DNA sequences recognized by RAG1 are the heptamer and nonamer of the RSS. To determine the extent of DNA binding of R121 and MBP-R1P7 to RSSs containing 12 or 23 bp spacers we utilized polyacrylamide gel mobility shift assays. R121 showed no evidence of binding,

whereas MBP-R1P7 bound weakly ( $K_d > 1 \mu\text{M}$ ; data not shown). However, the latter RSS DNA-protein complex could be easily displaced with non-specific competitor DNA, indicative of only weak non-specific DNA binding of MBP-R1P7.



**Figure 1.** Diagram illustrating the location of proposed domains within the primary sequence of RAG1 and the clones and sequences of expressed fragments of murine RAG1. A, The top bar represents the full-length RAG1 with the location of the metal-binding domains indicated by filled boxes. The location of potential nucleic acid binding regions are represented by hatched boxes. Lines indicate the positions of the R121 and core RAG1 domains. Beneath are bars representing the various clones produced of RAG1 placed in the corresponding position relative to the full-length protein. MBP refers to maltose-binding protein as each of these fragments was expressed as a chimeric protein. The name of the cloned proteins is to the left of its respective bar. B, Amino acid sequences of the C<sub>3</sub>HC<sub>4</sub>, ZFA, and ZFB regions in RAG1. The putative metal ligands for each sequence are shown coordinated to zinc.

**Table 1.** Metal stoichiometries of the cloned RAG1 fragments determined by atomic absorption spectroscopy

Protein <sup>a</sup>	Zn/protein molecule <sup>b</sup>
MBP-R163	3.1:1
MBP-R121*	3.2:1
R121	3.1:1
MBP-RF	1.7:1
MBP-R1P7	1.2:1
MBP	0.06:1

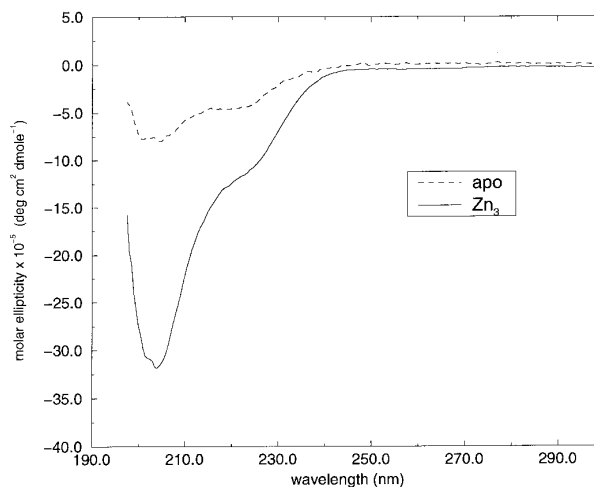
<sup>a</sup> The protein samples were in 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 5 mM 2-mercaptoethanol.

<sup>b</sup> Zinc concentrations were determined by fitting the atomic absorption of the protein samples to a standard curve determined for zinc standards from 1 to 15  $\mu$ M. The latter had a standard deviation of  $\pm 10\%$  throughout and the replicate protein sample determinations fell within the latter range.

### The C<sub>3</sub>HC<sub>4</sub> and ZFA motifs form a well-defined domain

MBP-R163 contains both the C<sub>3</sub>HC<sub>4</sub> and ZFA motifs in the RAG1 portion of the chimeric protein. MBP-R163 was found to be susceptible to partial proteolysis. During the standard purification 10 to 20% of the R163 domains were subject to proteolysis by endogenous *E. coli* proteases, which yielded a protein approximately 3.5 kDa smaller than the full-length MBP-R163. The proteolyzed form could be purified easily from the uncleaved chimeric protein. To determine the amino acid residues removed by the endogeneous proteolysis we separated the proteolyzed R163 domain from MBP. This was done by limited trypsin digestion, which results in a specific cleavage between the maltose binding protein and the RAG1 fragment. The amino acid composition of the resulting RAG1 fragment shows that proteolysis by the *E. coli* proteases removes Gly390 to Asp426, although the Gly389-Gly390 bond could be a specific cleavage site (Figure 1). As this site most likely represents a domain boundary, we cloned and overexpressed MBP-R121\*. Trypsin cleaves both MBP-R121 and MBP-R163 after Lys268 in the RAG1 sequence (five residues from the MBP/RAG1 junction). Thus, the remaining 121 residues (amino acid residues 269 to 389 of RAG1) containing both the C<sub>3</sub>HC<sub>4</sub> and ZFA sequences, referred to as R121, appear to form a structurally well-defined domain in RAG1. This conclusion is supported by the fact that proteolysis of larger fragments readily occurs to the boundaries defined by R121, but the remaining RAG1 polypeptide is resistant to further proteolysis. The MBP-R121\* chimeric protein actually contains 126 amino acid residues from RAG1 (264 to 389). For clarity (in comparison with R121) we refer to this protein as MBP-R121\* with the asterisk indicating five additional RAG1 residues at the N-terminal end of the RAG1 sequence.

The three zinc ions coordinated to R121 play a major role in maintaining the overall structural integrity of the protein as judged by the circular dichroism spectra (Figure 2). The three zinc holo-protein is predicted to contain 11%  $\alpha$ -helix and

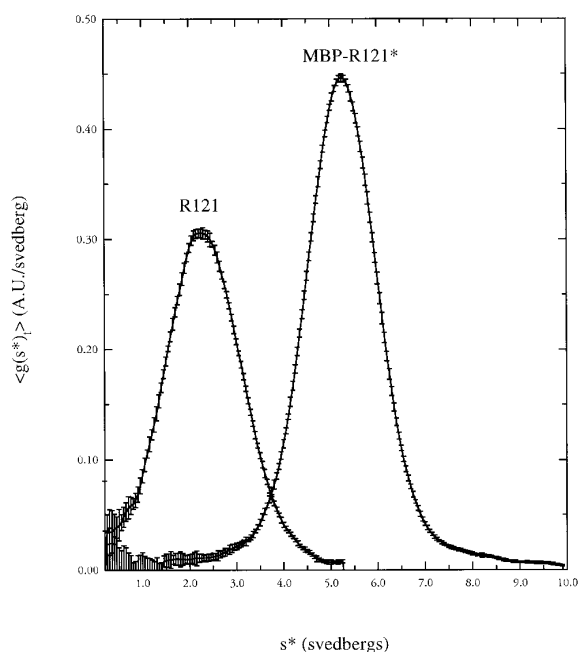


**Figure 2.** Circular dichroism spectra of R121 containing three zinc ions per protein molecule, and the apo form of R121 containing less than 0.2 zinc ion per protein molecule. The buffer used was 20 mM sodium phosphate (pH 6.0), 50 mM NaCl.

34%  $\beta$ -sheet based on the Sreerama-Woody self-consistent method for secondary structure analysis (Sreerama & Woody, 1993). Removal of the zinc ions produces a significant reduction of the molar ellipticity, suggesting a loss of most of the ordered secondary structure. This is not unexpected, since the energy associated with coordination of zinc ions is known to be necessary for the folding of several zinc-binding peptide motifs (Frankel *et al.*, 1987; Parraga *et al.*, 1988; Pan & Coleman, 1990). The favorable free energy associated with formation of the zinc complex must counteract the unfavorable entropy of folding. Similarly, the two zinc ions in the C<sub>3</sub>HC<sub>4</sub> motif may induce the formation of the three-stranded antiparallel  $\beta$ -sheet. In addition, the conformation of R121 seems to be metal specific as the cadmium form, in which the zinc ions are exchanged by non-dialyzable cadmium ions, results in a decrease in the solubility of the protein above 50  $\mu$ M. Cadmium has been observed to lead to significant changes in structural details at sites distant from the bound metal, even in a small monomeric protein (Gardner *et al.*, 1995).

### The R121 domain of RAG1 forms a specific dimer

The monomeric molecular masses of R121 and MBP-R121\* are calculated to be 13,870 and 56,768, respectively. However, elution profiles from gel filtration chromatography indicate solution molecular masses approximately twice these values for both proteins (data not shown). As the maltose-binding protein is known to be monomeric (Spurlino *et al.*, 1991), the dimerization of the chimeric MBP-R121\* protein appears to be induced by the R121 domain. In fact, the presence of the MBP subunits has proven advantageous in several



**Figure 3.** Analysis of the R121\* dimer by the  $dc/dt$  sedimentation velocity method. Shown is the apparent sedimentation coefficient distribution function,  $g(s^*)$  versus  $s^*$ , computed from the time derivative of the concentration profile using the method of Stafford (1992). The error bars represent the standard error of the mean for each point of  $g(s^*)$  propagated from the averaging of  $dc/dt$ . The protein sample concentrations were 1.23 mg/ml for R121 and 0.64 mg/ml for MBP-R121\*. The sample buffer was 20 mM sodium phosphate (pH 7.0), 150 mM NaCl, 5 mM 2-mercaptoethanol.

experiments discussed below. Thus, all measurements were done in parallel with R121 and MBP-R121\*.

To further establish that R121 is a dimer we used a combination of dynamic light-scattering and velocity sedimentation experiments to determine the molecular mass. Velocity sedimentation was used to determine the sedimentation coefficients by use of the time derivative method described in Materials and Methods. Apparent sedimentation coefficient distributions,  $g(s^*)$  as a function of  $s^*$ , for R121 and MBP-R121\* samples are shown in Figure 3. As the  $g(s^*)$  curves are symmetrical, we were able to obtain accurate values for the sedimentation coefficients. Dynamic light-scattering experiments were utilized to measure the translational diffusion coefficients for both R121 and MBP-R121\*. The same buffer conditions and protein samples were used in the determination of both the diffusion and sedimentation coefficients (20 mM sodium phosphate, 150 mM NaCl, 5 mM 2-mercaptoethanol). The resulting values for  $D_{20,w}^0$  and  $s_{20,w}^0$  were combined in the Svedberg equation to calculate the solution molecular masses (Table 2). The values verify that for both R-121 and MBP-R121\* the dimeric complex is the major species present in solution under the conditions tested.

Since protein associations may depend significantly on solution conditions, it was necessary to test the association state of the dimeric MBP-R121\* complex under various buffer conditions. The translational diffusion coefficients (measured by dynamic light-scattering) varied by only 5% over a pH range of 6 to 8 and NaCl concentrations of 25 mM to 1 M, consistent with the presence of a stable dimer of the MBP-R121\* chimeric protein over a wide range of solution conditions. All protein samples measured were at a concentration of 2 mg/ml. For all the samples the measured autocorrelation function fit well to a

**Table 2.** Physical characterizations summary for the R121 dimerization domain

Protein	R121	MBP-R121*	MBP2*
Calculated $M_r$			
Monomer	13,870	56,768	42,455
Dimer	27,740	113,537	
<b>A. Dynamic light-scattering and sedimentation velocity results</b>			
$s_{20,w}^0$ ( $\times 10^{13} \text{ s}^{-1}$ )	$2.44 \pm 0.02$	$5.88 \pm 0.02$	
$D_{20,w}^0$ ( $\times 10^7 \text{ cm}^2/\text{s}$ )	$7.97 \pm 0.10$	$5.35 \pm 0.10$	
$M_r$ (Svedberg eqn) <sup>a</sup>	29,232	104,943	
Frictional coefficient <sup>a</sup> ( $\times 10^8 \text{ g/s}$ )	5.08		
	(axial ratio $\sim 3.1$ ) <sup>a</sup>		
<b>B. Sedimentation equilibrium experiment</b>			
$M_r$	28,264	n.d.	
$K_d$ ( $\mu\text{M}$ )	n.d.	3.12	
<b>C. Small-angle X-ray scattering experiment</b>			
$M_r$	28,014	95,634	38,640
$R_g$ ( $\text{\AA}$ ) (Guinier analysis)	$21.6 \pm 0.5$	$40.7 \pm 5.7$	$22.3 \pm 0.8$
$R_g$ ( $\text{\AA}$ ) ( $P(r)$ function)	$23.4 \pm 0.1$	$45.1 \pm 0.2$	$22.5 \pm 0.1$
$I(0)/c^b$ ( $\times 10^3$ )	$2.9 \pm 0.3$	$9.9 \pm 1.1$	$4.0 \pm 0.4$
$d_{\text{max}}$ ( $\text{\AA}$ )	$89 \pm 1$	$159 \pm 6$	$69 \pm 1$

<sup>a</sup> Calculated as described in Materials and Methods.

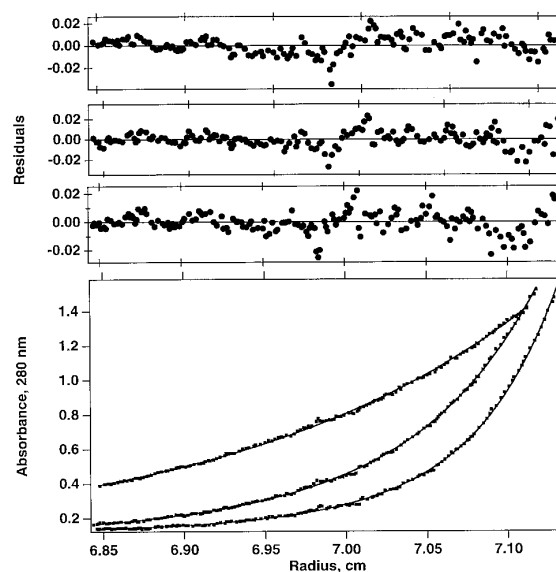
<sup>b</sup>  $I(0)/c$  for carbonic anhydrase =  $3.0 (\pm 0.3) \times 10^{-3}$  with a molecular mass of 28,980. The partial specific volumes for all the proteins were assumed to be 0.74 ml/g.

single exponential indicative of monodisperse samples.

The entire R121 domain is necessary to form a specific dimer, as the chimeric protein of MBP with the C<sub>3</sub>HC<sub>4</sub> subdomain alone aggregates extensively based on evidence from dynamic light-scattering. The  $D_{20,w}$  value of MBP-RF is approximately twofold lower than that measured for MBP-R121\*, and has a high polydispersity index (required to fit the time dependence of the light scattering). Removal of the ZFA sequence may expose core hydrophobic regions resulting in the aggregation observed, or alternatively ZFA may sterically prevent higher-order aggregation. The choice of boundaries for the C<sub>3</sub>HC<sub>4</sub> subdomain in the MBP-RF construct is similar to those used for the two C<sub>3</sub>HC<sub>4</sub> subdomains, one from EHV-gene 63 and one from PML, in which the structures have been determined by NMR (Barlow *et al.*, 1994; Borden *et al.*, 1995). Both of these ring fingers are monomeric in solution, in contrast to the zinc C<sub>3</sub>HC<sub>4</sub> subdomain in RAG1. The aggregation observed for MBP-RF is most likely due to the high representation of hydrophobic residues in the RAG1 C<sub>3</sub>HC<sub>4</sub> sequence, rather than the possibility that the boundaries chosen yield an unstable molecule. Specific dimer formation is dependent also on coordination of the zinc ions, as the apo MBP-R121\* chimeric protein is aggregated based on the measured  $D_{20,w}$  value. This value is approximately 2.5-fold lower, with a high polydispersity index, compared to the zinc-containing MBP-R121\* sample. As the overall structure of R121 is significantly unfolded upon removal of the zinc ions (Figure 2), it can be postulated that exposure of previously buried hydrophobic residues leads to non-specific aggregation. Reversal of the non-specific aggregation of the apoprotein with addition of zinc has proven unsuccessful under several buffer conditions.

### The R121 dimer of RAG1 forms with high affinity

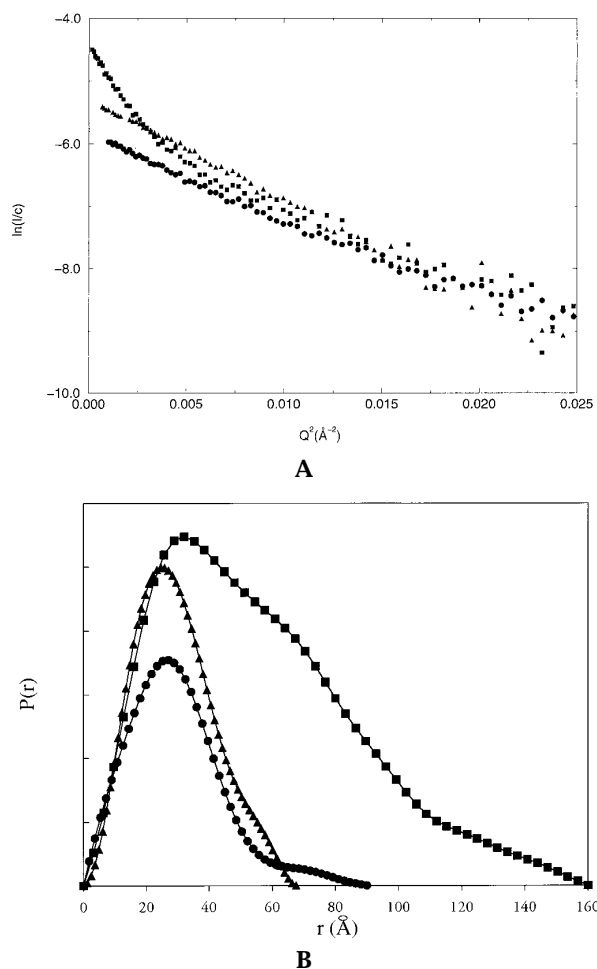
Determination of the energetics of formation of the R121 dimer provides insight into the magnitude of the forces involved in the dimerization of this unique domain. We utilized sedimentation equilibrium experiments to provide information on the thermodynamic parameters of dimer formation. With the larger extinction coefficient at 280 nm for the chimeric MBP-R121\* protein it was possible to work at low enough protein concentrations to detect significant concentrations of monomer. The concentration *versus* radial distance profiles for the MBP-R121\* chimeric protein at three speeds are shown in Figure 4. A global non-linear least-squares analysis demonstrates that the best fit to the data is an ideal, associating monomer-dimer model. In addition, a plot of the natural logarithm of the protein concentration *versus* the square of the radius for each of these data sets individually shows upward curvature of an associating system



**Figure 4.** Sedimentation equilibrium analysis of dimeric MBP-R121\*. MBP-R121\* was centrifuged in buffer containing 20 mM sodium phosphate (pH 7.0), 150 mM NaCl, 5 mM 2-mercaptoethanol. The sample (150  $\mu$ l) contained 0.70 mg of protein/ml. Equilibrium sedimentation data are shown for three rotor speeds: 8000 r.p.m., top curve; 12,300 r.p.m., middle curve; and 15,000 r.p.m., bottom curve (lower panel). The markers are the experimental data, and the continuous lines represent the global, least-squares fit to the data using an ideal monomer-dimer model. The residuals for each data set are shown in the three panels above: 8000 r.p.m. (top) to 15,000 r.p.m. (bottom).

(data not shown). The equilibrium constant was determined using the solution to the global fit by setting the monomer molecular mass to 56,768, and subsequently fitting for  $K_{abs}$ .  $K_{abs}$  was converted to a  $K_a$  (or  $K_d$ ) in the standard way using a monomer molar extinction coefficient of 67,300 at 280 nm (McRorie & Voelker, 1993). The value of the dissociation constant was calculated to be  $3.12 \times 10^{-6}$  M ( $\Delta G = 7.4$  kcal/mol) with 67% upper and lower confidence limits of  $3.70 \times 10^{-6}$  and  $2.60 \times 10^{-6}$  M, respectively.

We also performed sedimentation equilibrium experiments on the isolated R121 domain. Higher molar concentrations of R121 relative to MBP-R121\* were necessary for detection with the absorbance optics system, as R121 contains only three tyrosine residues in contrast to the 18 tyrosine and 8 tryptophan residues in MBP-R121\*. We observed a predominantly single species in solution consistent with R121 existing in the dimeric state. A global, least-squares analysis of equilibrium data collected at four speeds (16,300, 20,000, 24,500, and 30,000 rpm) demonstrates that the best fit to the data is a single ideal species with a  $M_r = 28,264$  with 67% upper and lower confidence limits of 27,824 and 28,704, respectively (data not shown). This solution molecular mass is within 2% of the predicted dimer molecular mass of 27,740 for the RAG1 fragment. No monomer was observed at this



**Figure 5.** Comparison of scattering data between R121 (●), MBP-R121\* (■), and MBP2\* (▲) as shown in A, Guinier plots; and B, distance distribution functions,  $P(r)$ , normalized for concentration.

concentration, as expected given the strong association observed for the chimeric protein.

### The shape of the R121 dimer is elongated

Small-angle X-ray scattering was used to examine the overall shape of the R121 dimerization domain. We represented the information contained in the scattering curves as distance distribution functions ( $P(r)$  curves), which is the real space equivalent of the scattering data (Figure 5). From the second moment of the  $P(r)$  function the radius of gyration ( $R_g$ ) of the particle can be determined. In addition, extrapolation of the scattered intensity to zero angle gives a value,  $I(0)/c$ , related to the molecular mass. Direct observation of the appropriate  $P(r)$  curve reveals the longest linear maximum dimension ( $d_{\max}$ ) of a particle providing information on overall shape. Results for measured  $R_g$ ,  $d_{\max}$ , and  $I(0)/c$  values of R121 are listed in Table 2.

The small-angle X-ray scattering measurements give an independent measure of the molecular mass of the isolated R121 domain. The  $I(0)/c$  value

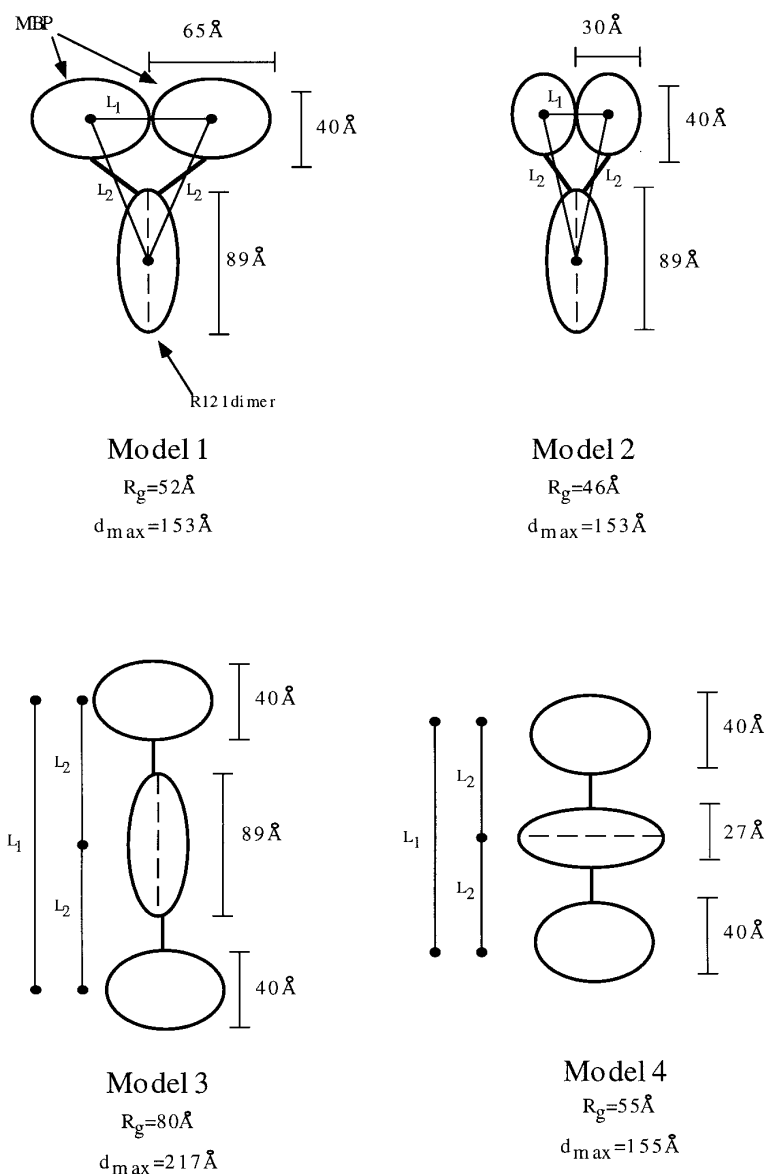
for R121 can be compared to the  $I(0)/c$  value for a carbonic anhydrase standard sample for which the molecular mass is known, and a molecular mass calculated for R121. The results confirm that R121 is a dimer (Table 2). Values determined for  $R_g$  and  $d_{\max}$  from the  $P(r)$  function provide the information concerning the overall shape of a molecule. An  $R_g$  value of 23.4  $\text{\AA}$  and a  $d_{\max}$  of 89  $\text{\AA}$  for the R121 dimer indicate that the complex forms an elongated structure relative to a sphere. If the overall shape of the dimer were spherical, the predicted radius of a sphere of equal volume would be 20  $\text{\AA}$  with a  $R_g$  value of 15.6  $\text{\AA}$  ( $R_g^2 = (3/5)r^2$ , for a sphere). As the  $d_{\max}$  value we determined for the R121 dimer is significantly larger than that predicted for the sphere (40  $\text{\AA}$ ), a simple approximation for the overall shape would be an elongated or prolate ellipsoid, although the  $P(r)$  function suggests a more complicated shape.

The  $R_g$  value for the R121 dimer was determined by two methods, using the Guinier approximation and the  $P(r)$  function (Table 2 and Figure 5). The value of  $R_g$  determined from the Guinier approximation is significantly smaller than the value obtained from the  $P(r)$  function. This may be a consequence of the elongated shape of R121 as it has been noted that the Guinier analyses for rod-shaped molecules lead to an underestimation of  $R_g$  (Glatter, 1982).

A uniform prolate ellipsoid with a maximum dimension of 89  $\text{\AA}$  and equivalent minor axes of 27  $\text{\AA}$  would enclose a volume consistent with the predicted molecular volume of the R121 dimer, given a molecular mass of 27,740 and a specific volume of 0.74  $\text{cm}^3/\text{g}$ . This uniform ellipsoid would have an axial ratio of 3.3. This estimate is consistent with the axial ratio of the R121 dimer determined using the measured sedimentation coefficient,  $s_{20,w}^0$ . From  $s_{20,w}^0$  the frictional coefficient was calculated yielding an axial ratio of 3.2 for R121 assuming a prolate ellipsoid (see Materials and Methods).

### The monomeric subunits within the R121 dimer are in a parallel orientation

Since the R121 dimer has an elongated shape the monomeric subunits may assume either a parallel or an antiparallel orientation. To distinguish between these two possible orientations we were able to use the chimeric MBP-R121\* dimer to our advantage, as the relative positions of the two MBP units within the chimeric MBP-R121\* dimer would indicate the alignments of the R121 monomeric subunits within the dimerization domain. We collected scattering data for the MBP-R121\* dimer as well as its component domains (the R121 dimer and the isolated monomeric MBP2\* subunits). MBP2\* includes MBP plus the C-terminal 16 residue linker. Confirmation that the components are in the appropriate aggregation states is provided by a simple mathematical combination of the extrapolated  $I(0)/c$  values (Table 2).



**Figure 6.** Schematics of four models for subunit orientation in the R121 dimer. The  $R_g$  values (calculated according to equation (2) in the text) and the  $d_{\max}$  values (obtained directly from the models) are shown below the corresponding model. For these calculations the distance between the MBP subunits and the R121 dimer was taken to be 24 Å. For model 1,  $L_1 = 65 \text{ \AA}$  and  $L_2 = 94.3 \text{ \AA}$ . For model 2,  $L_1 = 30 \text{ \AA}$  and  $L_2 = 89.8 \text{ \AA}$ . The distance for  $L_2$  in models 1 and 2 is derived by bisecting the  $L_1$ - $L_2$ - $L_2$  triangle through the  $L_2$ - $L_2$  angle. As the distance of this bisect and  $L_1$  are known,  $L_2$  can be calculated. Specifically for models 1 and 2,  $L_2^2 = (L_1/2)^2 + (40/2 + 24 + 89/2)^2$ . For model 3,  $L_1 = 177 \text{ \AA}$  and  $L_2 = 88.5 \text{ \AA}$ ; and for model 4,  $L_1 = 115 \text{ \AA}$  and  $L_2 = 57.5 \text{ \AA}$ . Model 4 is the best representation for the MBP units attached to the R121 dimer along the long axis. There are many possibilities as to where the MBP units may attach along the long axis of the R121 dimer. However, out of these different possibilities model 4 would result in the smallest  $R_g$  value, which is still significantly larger than the experimental result. In addition, the dimeric interface between the two R121 monomeric subunits could be along the minor axis in model 4 rather than along the major axis as is necessary in models 1 to 3.

Specifically, the  $I(0)/c$  value for the chimeric protein is equivalent to the sum of the  $I(0)/c$  for the R121 dimer plus twice the  $I(0)/c$  value for MBP2\* (within 11%). This internal reference illustrates that the chimeric protein and the isolated R121 domain are dimeric and that MBP2\* is monomeric in solution. Molecular masses determined by comparison to a carbonic anhydrase standard supports this conclusion as well (Table 2). The measured  $R_g$  and  $d_{\max}$  values for these proteins are listed in Table 2. Again we see a difference in the  $R_g$  values determined from the  $P(r)$  function and the Guinier plot for the chimeric protein, possibly resulting from the elongated shape of the MBP-R121\* dimer. The  $R_g$  values for MBP2\* compare well between the two different analyses.

We developed four models for the possible orientations of the subunits in the chimeric dimer (Figure 6). Models 1 and 2 consist of the R121 monomers orientated in a parallel fashion (MBP

monomers on same end of the R121 dimer) and model 3 consists of the subunits antiparallel (MBP monomers on opposite ends of the R121 dimer), while model 4 can be in either orientation. The two parallel models are distinguished by the relative orientation of the MBP units to each other. In model 4 the MBP units are attached near the midpoint of the long axis, in contrast to models 1 to 3 where the attachment point is at the small end(s) of the R121 dimer. We evaluated the validity of our models by using the parallel axis theorem (Engelman & Moore, 1975). In general, this theorem is based on the premise that two subunits of an aggregated particle can be equated by:

$$R_g^2 = \omega_1 R_{g1}^2 + \omega_2 R_{g2}^2 + \omega_1 \omega_2 L^2 \quad (1)$$

where  $\omega_1$  and  $\omega_2$  are the weight fractions and  $R_{g1}$  and  $R_{g2}$  are the radii of gyration of the two component parts.  $L$  is the distance between the center of masses of the two parts and  $R_g$  is the



radius of gyration of the complete particle. In our particular instance we have three separable components of the chimeric protein, with two MBP2\* units and the R121 dimer. Hence the equation becomes:

$$R_g^2 = 2\omega_1 R_{g1}^2 + \omega_2 R_{g2}^2 + \omega_1^2 L_1^2 + 2\omega_1 \omega_2 L_2^2 \quad (2)$$

where  $R_g$  is the radius of gyration of the MBP-R121\* dimer,  $R_{g1}$  and  $\omega_1$  are the radius of gyration and weight fraction of MBP2\* monomer, and  $R_{g2}$  and  $\omega_2$  are the radius of gyration and weight fraction of the R121 dimer.  $L_1$  is the distance between the centers of mass of the MBP units and  $L_2$  is the distance between the centers of mass of the R121 dimer and each MBP unit, derived as shown in Figure 6. In calculating the parameters for each model, the  $R_g$  values determined from the  $P(r)$  functions were used in equation (2). In addition, we modeled each of the components within the models as triaxial ellipsoids. The overall shape of the R121 dimer is considered to approximate a prolate ellipsoid with a major axis of 89 Å ( $d_{\max}$  from the  $P(r)$  curve) and minor axes estimated to be 27 Å. The dimensions of MBP (65 Å × 40 Å × 30 Å) were obtained from the crystal structure (Spurlino *et al.*, 1991). In addition, the relative position of the carboxy terminus in the crystal structure was used in modeling the extension of the linker from MBP. This linker between MBP and the R121 domain is assumed to be in an  $\alpha$ -helical conformation giving a 24 Å length for the 16 amino acid linker. From the results outlined in Figure 6, it is clear that model 2 with a calculated  $R_g$  of 46 Å and  $d_{\max}$  of 153 Å agrees exceptionally well with the measured values for MBP-R121\* ( $R_g = 45.1$  Å and  $d_{\max} = 159$  Å). The best-fitting model can be summarized as having a parallel orientation of the R121 subunits with the MBP units positioned side-by-side along their long axes. Clearly the model with an antiparallel orientation is not in agreement with our measured results. The calculated values of  $R_g$  and  $d_{\max}$  seen for model 3 in Figure 6 are very different from the experimentally determined values for the chimeric protein. Similarly the  $R_g$  calculated for model 4 is 10 Å larger than the measured  $R_g$  for MBP-R121\*. Even though model 4 results in a  $d_{\max}$  value close to the measured value, the  $s_{20,w}^0$  value for MBP-R121\* (Table 2) generates a frictional coefficient more in agreement with the significantly larger axial ratio expected from model 2. An inexact factor in the modelling shown in Figure 6 is the distance between MBP and the R121 domain determined by the linker, which is assigned as 24 Å. However, by varying this distance over a wide range (down to 12 Å) model 2 still gives the closest agreement to the experimental results, with the other models differing significantly in every case.

## Discussion

RAG1, along with RAG2, has been identified as an essential component in the mechanism of V(D)J recombination (Schatz *et al.*, 1989; Oettinger *et al.*,

1990). It was found by deletional mutagenesis that only the carboxyl-terminal two-thirds of RAG1 is necessary for recombination *in vivo* of a plasmid substrate (Sadofsky *et al.*, 1993; Silver *et al.*, 1993). This region, referred to as core RAG1, consists of amino acid residues 384 to 1008 (Figure 1). Core RAG1 is particularly sensitive to mutations as nine out of ten small deletions within this sequence prove deleterious to V(D)J recombination (Sadofsky *et al.*, 1993). With the recent report that purified core RAG1 and a core RAG2 perform site-specific cleavage of a DNA fragment containing an RSS, the current proposal is that the core of RAG1 participates directly in the recognition and cleavage process (McBlane *et al.*, 1995).

The current assay used to investigate the efficiency of V(D)J recombination *in vivo* relies on recombination of a transfected, extrachromosomal plasmid substrate in which the RSSs are separated by a short distance (typically less than 1 kb). Although the core RAG proteins are sufficient for recombination in this assay, recombination of the endogenous antigen receptor loci may be more complex, for at least two reasons. First, endogenous V(D)J recombination events often span large genomic distances (hundreds of kb), suggesting the need for a mechanism to facilitate the interaction between the recombining gene segments. And second, assembly of antigen receptor genes by V(D)J recombination is a carefully regulated developmental process. Regulation is imposed by poorly understood mechanisms which control access of the recombination machinery to the substrate DNA, suggesting that the recombination machinery is able to sense aspects of chromatin structure, as demonstrated to be the case for methylation (Hsieh & Lieber, 1992). Furthermore, it remains unknown whether V(D)J recombination occurs preferentially or exclusively in subdomains of the nucleus. Therefore, functions provided by the N-terminal third of the RAG-1 protein may contribute to the proper organization and/or targeting of the recombination complex and thereby facilitate endogenous gene rearrangement, but such functions might not be readily apparent in the assays used thus far. In fact, recent experiments have demonstrated that core RAG1 is suboptimal for catalysis of V(D)J recombination even in the standard *in vivo* recombination assay. Nearly full length RAG-1 (amino acid residues 1 to 1008) is substantially more active than core RAG1, despite the fact that core RAG1 is typically more highly expressed. Only amino acid residues 216 to 1008 are needed to observe the increase in activity over that seen with the core (C. McMahan, N. Rao & D.G.S., unpublished results). This indicates that at least part of the N-terminal third of RAG1 enhances the activity of the core region during the basic catalytic steps of recombination in the *in vivo* assay.

Once the primary amino acid sequence of RAG1 was known, it was found that a sequence within the N-terminal third of the protein was homologous to a zinc-binding motif referred to as the  $C_3HC_4$  or the

zinc ring finger domain (Schatz *et al.*, 1989). We have found that the  $C_3HC_4$  motif in RAG1 binds two zinc ions, as expected. In addition we have discovered two additional zinc-binding sequences in RAG1 whose sequences are similar to the  $C_2H_2$  zinc finger motif. Each of these motifs binds one zinc ion. We refer to the two zinc fingers as ZFA and ZFB. The relative positions of these subdomains are shown in Figure 1A, with the primary sequences given in Figure 1B. The amino acid sequences of ZFA and ZFB deviate somewhat from the sequence and spacing of the classical zinc finger, Y/F-X-C-X<sub>2-4</sub>-C-X<sub>3</sub>-F-X<sub>5</sub>-L-X<sub>2</sub>-H-X<sub>3-5</sub>-H (Berg, 1990; Coleman, 1992). While both ZFA and ZFB sequences fit with the conserved short ligand spacing of C-X<sub>2-4</sub>-C and H-X<sub>3-5</sub>-H for the zinc finger motif, neither contain the conserved F and L residues usually observed in the long spacer between the second and third ligands. However, ZFA and ZFB contain a number of hydrophobic side-chains in this region, including leucine residues, sufficient to form a hydrophobic core for a zinc finger. In addition, many zinc finger structures have been determined by NMR and X-ray crystallography, with several of these having significantly deviant sequences (Jasanoff *et al.*, 1992; Omichinski *et al.*, 1992; Qian & Weiss, 1992; Pavletich & Pabo, 1993). Despite these differences, these "non-classical" zinc fingers display the same general structural characteristics consistent with the zinc fingers found in TFIIIA. In fact, the sequence of ZFB is ~40% identical (~60% homologous) to the sequence of the 13th zinc finger of Xfin and the third zinc finger of ZG64.

As seen in Figure 1A, ZFB is located within the essential core RAG1 sequence. Whether ZFB plays a crucial role in V(D)J recombination is unclear. We have found that a fragment of RAG1 containing ZFB displays only weak and apparently non-specific affinity for DNA. However, one of the deletion mutants of core RAG1 which abolishes V(D)J recombination occurs within the spacer region between the second and third ligands of ZFB, replacing residues 735 to 740 with a valine and an aspartate giving a four residue deletion (Sadofsky *et al.*, 1993). This mutation would most certainly hinder zinc binding in this subdomain, and suggests an important role for ZFB in the activity of core RAG1. Conservative mutations of the putative ligands within ZFB would provide a more definitive result.

ZFA is positioned 19 amino acid residues after the carboxy terminus of the  $C_3HC_4$  motif. Based on proteolysis studies both of these zinc-binding subdomains are included in a structurally well-defined domain, referred to as R121 (Figure 1). Using a combination of biophysical techniques we have found that this domain forms a specific dimer with a relatively high energy of formation. As the  $C_3HC_4$  motif is found in a variety of proteins with varying cellular functions, the physiological role of this zinc-binding motif has been of increasing interest. Proposals have centered around roles in

DNA binding or protein-protein interactions. To date there have been no reports of  $C_3HC_4$  subdomains recognizing specific DNA sequences. In fact, we have found that the R121 domain does not bind to the recombination signal sequences based on gel mobility shift assays. It has been suggested that the  $C_3HC_4$  motifs in EHV1 gene 63, PML, and TRAF2 (TNF receptor-associated factor) are involved in heterologous protein-protein interactions, although the targets for these associations have not been identified (Borden *et al.*, 1995; Rothe *et al.*, 1995). The evidence presented here directly demonstrates that the  $C_3HC_4$  motif of RAG1 participates in protein-protein interactions via homodimer formation.

However, the  $C_3HC_4$  motif of RAG1 alone is insufficient for specific dimer formation, as removal of ZFA results in extensive aggregation of the  $C_3HC_4$  subdomain. Thus, we define the R121 dimerization domain of RAG1 as a multiclass zinc-binding domain, since both a zinc-coordinated  $C_3HC_4$  motif and a zinc finger are required. This unique arrangement of different classes of zinc-coordinating motifs acting in conjunction may be a feature seen among other  $C_3HC_4$ -containing proteins. For example, several of these proteins contain both a  $C_3HC_4$  motif and a zinc-coordinating B-box within 32 to 37 amino acid residues (Freemont, 1993). A second example includes TRAF2 and CRAF1 (CD40 receptor associated factor), in which both proteins contain multiple putative zinc fingers following the carboxy terminus of the  $C_3HC_4$  motif (Cheng *et al.*, 1995). Additionally, this domain of RAG1 is an example of a zinc-dependent dimerization domain whose function is not induced by DNA binding in contrast to the zinc-containing dimerization domain of the glucocorticoid receptor (Luisi *et al.*, 1991). A second example is the LIM domain which has been found to be involved in both heterologous protein-protein associations and homodimer formation (Feuerstein *et al.*, 1994; Schmeichel & Beckerle, 1994; Bach *et al.*, 1995). The LIM domain also coordinates two zinc ions, but with no apparent structural homologies to the  $C_3HC_4$  motif (Perez-Alvarado *et al.*, 1994; Schwabe & Klug, 1994).

Small-angle X-ray scattering results provided further insight into the overall shape and alignment of subunits within the R121 dimer. The interpretation of the  $R_g$  and  $d_{max}$  values, in which the overall shape of the dimerization domain can be described best as an elongated uniform prolate ellipsoid, leads to a model with a parallel orientation of the monomeric subunits. Results from recent characterizations of murine RAG1 expressed in the murine B lymphoma cell line M12 indicate dimerization of the full-length protein (Schatz & Leu, 1996). Thus, from the small-angle X-ray scattering results we can assume that the topology of the full-length dimeric protein would include the monomers in a parallel orientation. Such an arrangement may result in close proximity of the catalytic domains from each subunit, which may be necessary for efficient

activity of the enzyme. Further biochemical and structural characterizations of RAG1 will test the validity of these proposals.

The R121 domain of RAG1 is in an ideal location within the primary sequence to serve as the dimerization domain for the full-length protein. As seen in Figure 1, R121 is located immediately N-terminal to the core RAG1 region. In such a position, the R121 dimerization domain could orient the core RAG1 subunits for enhanced protein-protein or protein-DNA interactions. Focusing on the immediate region of the dimerization domain, R121 is flanked on both sides by sequences with a net positive charge (Figure 1). Bordering on the N-terminal side is a region (amino acid residues 216 to 270) containing two stretches of basic amino acid residues which have been shown to be important for the interaction between RAG1 and SRP1 (Cortes *et al.*, 1994; Spanopoulou *et al.*, 1995). SRP1 is a member of a family of proteins with a role in nuclear transport (Gorlich *et al.*, 1994) and appears to affect the intranuclear localization of RAG1 (Spanopoulou *et al.*, 1995). These sequences, in combination with R121, constitute the additional region that when appended to core RAG1 to give amino acid residues 216 to 1008 results in the highest recombination activity in the *in vivo* assay (C. McMahan, N. Rao & D.G.S., unpublished results). Also, flanking the C-terminal side of R121 is a 70 residue sequence that we have shown to contain a high percentage of  $\alpha$ -helical structure and has DNA binding capabilities as an isolated domain (K.K.R., D.G.S. & J.E.C., unpublished results). This latter region is within the core RAG1 sequence and has proven to be indispensable for V(D)J recombination.

In summary, a specific dimerization domain in RAG1 includes the presence of two different classes of zinc-binding sequences, a  $C_3HC_4$  motif and a zinc finger. This finding of a protein domain which contains multiple classes of zinc binding motifs reflects the additional structural and functional diversity that can be created by the combination of these small stable subdomains.

## Materials and Methods

### Cloning and purification

Murine RAG1 fragments were cloned as chimeric proteins with the maltose-binding protein (MBP). The MBP fusions were constructed in vector pMALc2 (New England Biolabs; Beverly, MA). Various RAG1 DNA fragments were amplified using the polymerase chain reaction (PCR) from pRAG1, a vector encoding the full-length sequence of RAG1. The vector encoding MBP-RF includes amino acid residues 277 to 339 in the RAG1 sequence. The primers used to construct this fragment were; 5'-CTCAGTACCAAGGTTAACGC-CGTG-3' and 5'-CAGAAAGGGATCCACTCAGCTCTC-3'. MBP-R163 (amino acid residues 264 to 426 of RAG1) was constructed using primers 5'-CCTGAAGAA-GATCGTAACTGC-3' and 5'-GTCAAGCAGACG-GATCCTCAATCTC-3'. MBP-R121\* (amino acid residues

264 to 389 of RAG1) was constructed using primers 5'-CCTGAAGAAGATCGTAACTGC-3' and 5'-AGAT-GCTGGCGAGATCTTCACCCTT-3'. The purified PCR products were digested with *HpaI* and *BamHI* (or *BglIII* for the R121\* PCR fragment) and subsequently ligated into pMALc2 that had been digested with *XmnI* and *BamHI*. MBP-R1P7 (amino acid residues 590 to 758) was constructed using primers 5'-CGGGATATCCTGGACC-CCTTCACAGTGGTGGTAAAG-3' and 5'-CGCGAA-TTCTAGGATCCCTCATAGCGCTGCAGGTTTC-3'. The purified PCR product was digested with *EcoRV* and *BamHI* and ligated into the *XmnI*-*BamHI* sites of a pMALc2 vector which had been modified to encode a C-terminal extension of six histidine residues.

All of the clones were transformed into *Escherichia coli* BL21 cells. The bacteria were grown at 25°C in Luria Broth with 100  $\mu$ g/ml ampicillin. Protein expression was induced by addition of isopropyl- $\beta$ -D-thiogalactopyranoside to a concentration of 0.2 mM. After 12 to 14 hours the bacteria were harvested by centrifugation and the pellets frozen.

MBP-RF, MBP-R163, and MBP-R121\* were all purified with a similar procedure. Frozen pellets were thawed on ice; resuspended in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 0.1 mM ZnCl<sub>2</sub>, 1 mM EDTA, 1 mM dithiothreitol, and lysed by sonication. Insoluble material remaining after lysis was removed by centrifugation at 12,000 rpm for 30 minutes. Poly(ethylenimine) was added to the supernatant to a concentration of 0.3%, and the sample incubated on ice for 20 minutes. After centrifugation as described above, the supernatant was dialyzed against purification buffer (20 mM Tris-HCl (pH 8.0), 0.1 mM ZnCl<sub>2</sub>, 5 mM 2-mercaptoethanol) plus 25 mM NaCl at 4°C. (SDS-PAGE at this point in the procedure showed that a major proteolytic product of MBP-R163 formed during cell lysis with a molecular mass approximately 3.5 kDa less than that of the full-length protein.) The dialysate was loaded onto a 50 ml Q-Sepharose column and a gradient run from 25 mM to 300 mM NaCl. Full-length MBP-R163 flowed through the column. The MBP-R163 proteolyzed protein, MBP-R121\*, and MBP-RF chimeric proteins eluted from the column between 150 and 200 mM NaCl concentrations. Fractions containing the MBP chimeric proteins, as identified by SDS-PAGE, were further purified by affinity chromatography using a column of cross-linked amylose resin pre-equilibrated with purification buffer plus 200 mM NaCl. The chimeric protein was eluted with buffer containing 10 mM maltose, concentrated to ~2 to 3 ml, and loaded onto a P-100 gel filtration column to remove small amounts of MBP produced by proteolysis. Fractions containing the chimeric protein, judged >95% pure by Coomassie-stained polyacrylamide gels, were collected and pooled.

The purification of MBP-R1P7 was similar up to the poly(ethylenimine) precipitation step. The supernatant after centrifugation (as described above) was loaded directly onto the amylose column. The protein was eluted as described above. Fractions containing the protein were then combined and dialyzed against 20 mM Tris-HCl (pH 8.0), 5 mM imidazole, 500 mM NaCl. The dialyzed protein was subsequently loaded onto a His-Tag metal chelation column (Novagen, Madison, WI). The column was washed with 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 60 mM imidazole, and the protein eluted with 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 500 mM imidazole. Fractions containing MBP-R1P7 were judged >95% pure by Coomassie-stained polyacrylamide gels. Purified MBP2\* was kindly provided by Matthew Junker.

Cleavage of the R121 fragment from the maltose binding protein was achieved by limited trypsin proteolysis. Samples with weight ratios of 50 to 100:1 chimeric protein to trypsin were incubated at room temperature for two hours before loading onto a 50 ml Q-Sepharose column pre-equilibrated with purification buffer plus 25 mM NaCl. This reaction resulted in >95% cleavage of the chimeric protein between the MBP and the RAG1 fragment. Protein was eluted with a gradient from 25 mM to 200 mM NaCl. R121 eluted before MBP at NaCl concentrations of 70 to 130 mM. Fractions containing R121 could be visualized by SDS-PAGE and silver staining. To ensure separation from trace amounts of trypsin, fractions containing R121 were concentrated to ~2 ml and loaded onto a P-30 gel filtration column. Fractions containing R121, judged >95% pure by silver-stained polyacrylamide gels, were collected and pooled.

In order to determine amino acid composition and to aid in determination of extinction coefficients, amino acid analyses were done routinely. The amino acid analyses were done at the W. M. Keck Foundation Biotechnology Resource Laboratory at the Yale University School of Medicine. Extinction coefficients determined for R121 and MBP-R121\* were  $5140 \text{ M}^{-1} \text{ cm}^{-1}$  at 275 nm and  $67,300 \text{ M}^{-1} \text{ cm}^{-1}$  at 280 nm, respectively.

### Metal exchange and analysis

Atomic absorption spectroscopy was done using an Instrumentation Laboratory IL157 spectrometer (Lexington, Massachusetts). The metal contents of the RAG1 proteins were measured on solutions of 3 to 10  $\mu\text{M}$ . These concentrations corresponded to the calibration curve determined prior to each protein sample measurement using 1 to 15  $\mu\text{M}$  Zn or Cd standard solutions.

Zinc ions in the RAG1 proteins were exchanged for cadmium by dialysis against nitrogen-saturated 50 mM Tris-acetate (pH 6.0), 200 mM NaCl, 2 mM 2-mercaptoethanol, and a fivefold excess of cadmium to protein at 4°C for 17 hours. The protein was then dialyzed against 20 mM phosphate (pH 7.0), 200 mM NaCl, 2 mM 2-mercaptoethanol at 4°C. Atomic absorption spectroscopy was done to determine the relative number of cadmium ions bound. At this point the number of cadmium ions bound was generally >2.5 and the number of zinc ions bound <0.3 per protein molecule. R121 and MBP-R121\* apoproteins were obtained by extensive dialysis against 50 mM Tris-acetate (pH 6.0), 200 mM NaCl, 5 mM 2-mercaptoethanol, 10 mM EDTA. The amount of zinc was determined by atomic absorption spectroscopy. The above dialysis was repeated until the ratio of zinc/protein molecule was less than 0.2.

### Circular dichroism

Circular dichroism spectra were collected on an Aviv model 62DS CD spectropolarimeter (AVIV Instruments, Lakewood, New Jersey) using a Hellma cuvette with a 0.2 mm path length. Proteins samples were dialyzed against buffer containing 20 mM sodium phosphate (pH 6.0), 50 mM NaCl. The final spectrum of each sample is the average of five separate spectra with a step size of 0.5 nm, one second signal integration time, and 1.5 nm bandwidth. The spectra were baseline-corrected using the dialysis buffer. The temperature during the scans was maintained at 25°C with a water-jacketed cuvette holder. The molar ellipticity was determined by using protein concentrations obtained from amino acid analysis.

Analysis of the circular dichroism spectra in estimating the secondary structure content was done using the Fortran program SELCON kindly provided by Sreerama and Woody. The method used in the secondary structure determination has been described (Sreerama & Woody, 1993).

### Gel electrophoretic mobility-shift assays

Gel electrophoretic mobility-shift assays were done using synthetic oligonucleotides containing the conserved recombination signal sequence with either 12 or 23 bp spacers between the conserved heptamer and nonamer sequences. The single-stranded oligonucleotides were annealed to yield the complementary duplex DNA. The DNA fragments were  $^{32}\text{P}$ -labeled with [ $\gamma$ - $^{32}\text{P}$ ]ATP using phage T4 polynucleotide kinase. Gel mobility shift assays were done as described with the exception that poly(dIdC) was used as the competitor DNA (Rodgers & Coleman, 1994).

### Standardized gel filtration chromatography

The elution profile of R121 and MBP-R121 from gel filtration chromatography was examined using a Pharmacia Superdex 75 column for R121 and a Pharmacia Superdex 200 column for MBP-R121. Protein standards used varied from molecular masses of 17 kDa to 120 kDa. The column buffers consisted of 20 mM phosphate (pH 7.0), 200 mM NaCl, 5 mM 2-mercaptoethanol. Protein elution was monitored by optical absorbance at 280 nm.

### Sedimentation equilibrium

The solution molecular mass and the equilibrium constant were determined by sedimentation equilibrium using a Beckman Optima XL-A analytical ultracentrifuge. Experiments were carried out on a 150  $\mu\text{l}$  sample in an An-60ti rotor using double sector centerpieces. Data from multiple speeds were collected at intervals of 0.001 cm using tenfold averaging at 20°C. Typically, absorbance versus radius scans were collected at 90 minute intervals following an eight hour delay after setting the rotor speed. Successive scans were compared graphically using the XL-A software to ensure that the sample reached equilibrium. Data were initially analyzed individually for self consistency on a PowerMac 7100 using IGOR PRO software and the Power PC XLA Ref. Std 1.01 kindly provided by Preston Hensley at SmithKline Beecham Pharmaceuticals. The final parameters were determined by a global non-linear least-squares fit of three or more data sets using the Macintosh version NONLIN algorithm described by Johnson *et al.* (1981). Goodness of fit was determined by examination of the residuals and minimization of the variance. The partial specific volume of both proteins was calculated based on the amino acid composition as determined by the sequence translated from the expression plasmid (Cohn & Edsall, 1943). These values were 0.74 and 0.7406 for the MBP-R121\* and R121 proteins, respectively. The solvent density was estimated to be 1.006.

### Sedimentation velocity

Sedimentation coefficients were determined by sedimentation velocity using a Beckman Optima XL-A analytical ultracentrifuge. Experiments were carried out

in an An-60ti rotor using double-sector centerpieces with solute concentrations ranging from 0.5 to 2.5 mg/ml at 55,000 rpm at 20°C. Sedimentation coefficient distributions ( $g(s^*)$ ) were calculated from the time derivative of the concentration distributions ( $dc/dt$ ) using the method of Stafford (1992). The weight average sedimentation coefficient,  $s_{wt}$ , was estimated from the  $g(s^*)$  function by the following equation:

$$s_{wt} = \frac{\int g(s^*)s^*ds^*}{\int g(s^*)ds^*} \quad (3)$$

Extrapolations of  $s_{20,solvent}$  to  $s_{20,w}$  were made from a least-square fit of data for  $s$  against concentration by regression analysis. Values of  $s_{20,w}^0$  were determined from  $s_{20,solvent}^0$  as described by Laue *et al.* (1992).

### Dynamic light-scattering

Translational diffusion coefficients ( $D_t$ ) were measured using a DynaPro-801 dynamic light-scattering instrument. Light-scattering data were obtained using an incident beam with a wavelength of 788 nm collected at angles of 90°. The temperature of the sample cell was monitored during data collection. The diffusion coefficient is calculated using an autocorrelation function. Along with  $D_t$  a polydispersity index is obtained. This index describes the goodness of fit of the autocorrelation function to a single exponential decay. Protein samples were filtered through Whatman Anatotop 10 filters (0.02 or 0.1  $\mu$ m) immediately prior to measurement.

The  $D_{t,solvent}^0$  values were determined by measuring a concentration series of each protein and extrapolating  $D_{t,solvent}$  to zero protein concentration. The values were then corrected to standard conditions of water at 20°C to give  $D_{20,w}^0$ .

### Calculation of molecular mass and axial ratio from $s_{20,w}^0$ and $D_{20,w}^0$

The values determined for  $s_{20,w}^0$  and  $D_{20,w}^0$  were used to calculate the molecular mass using the Svedberg equation:

$$M = \frac{s_{20,w}^0 RT}{D_{20,w}^0 (1 - \bar{v}_2 \rho)} \quad (4)$$

where  $\bar{v}_2$  is the specific volume of the protein and  $\rho$  is the density of pure water at 20°C.

The frictional coefficient ( $f_{20,w}^0$ ) was obtained using the measured  $s_{20,w}^0$  value in the following equation:

$$f_{20,w}^0 = \frac{M_f (1 - \bar{v}_2 \rho)}{N_0 s_{20,w}^0} \quad (5)$$

where  $M_f$  is the molecular mass and  $N_0$  is Avogadro's number. For the case of R121, the molecular mass of the dimer calculated from the amino acid sequence and including three zinc ions per monomer is 27,740. The frictional coefficient ( $f_0$ ) of an equivalent hydrated sphere is calculated from the relationship:

$$f_0 = 6\pi\eta [3M(\bar{v}_2 + \delta\bar{v}_1)/4\pi N_0]^{1/3} \quad (6)$$

where  $\eta$  is the viscosity of water at 20°C,  $\bar{v}_1$  is the specific volume of pure water, and  $\delta$  is the degree of hydration. For R121 we calculated  $\delta$  to be 0.35 g H<sub>2</sub>O/g protein according to Kuntz (1971). The frictional ratio,  $f_{20,w}^0/f_0$ , was used to calculate the maximum axial ratio,  $a/b$ , of a prolate ellipsoid using the method of Laue *et al.* (1992), where  $a$  is the major semi-axis and  $b$  is the minor semi-axis.

### Small-angle X-ray scattering

The protein samples were dialyzed against 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 5 mM 2-mercaptoethanol, 0.1 mM maltose at 4°C overnight. As substrate binding results in a conformational change in MBP, we added excess maltose to the dialysis buffer to ensure that MBP was in the substrate-bound form. Details of the small angle X-ray scattering apparatus will be described elsewhere (Z. Bu, H. H. Wyckoff & D.M.E., unpublished results). In brief, the X-ray source was a Rigaku RU-300 rotating-anode generator operated at 50 kV and 180 mA producing 1.54 Å CuK<sub>α</sub> radiation. The beam was point collimated with an incident beam diameter of 1.2 mm. A 2.3 m helium flushed path was placed between the sample holder and the detector to reduce the scattering from air. The detector was a two-dimensional multiwire detector with a sensitive area of 290 mm × 256 mm. For the chimeric MBP-R121\* dimer the measured  $Q$  range was from 0.014 to 0.36 Å<sup>-1</sup>. All scattering intensities were put on an absolute scale by measuring the incident beam intensity with a device to be described (Z. Bu, H. H. Wyckoff, D.M.E., unpublished results). The rest of the instrumental factors remained unchanged during the measurements. Scattering data were collected in two to six hours with the counting error less than 1% in the smallest angles. All measurements were at room temperature. In calculation of the net scattering from the protein molecules, scattering from the final dialysate was used for buffer subtraction.

In small-angle X-ray scattering experiments the scattered intensity  $I(Q)$  is measured as a function of the length of the scattering vector  $Q = 4\pi \sin\theta/\lambda$ , where  $\lambda$  is the wavelength of the radiation and  $2\theta$  is the scattering angle. We analyzed the scattering data using the Guinier analysis and the distance distribution function,  $P(r)$ . From the Guinier analysis an estimate of the radius of gyration ( $R_g$ ) was obtained. The  $P(r)$  functions were calculated using the real-space algorithms (Semenyuk & Svergun, 1991) with the program GNOM. The  $P(r)$  function is related to the probability distribution of interatomic vectors within an individual scattering particle. The maximum linear dimension,  $d_{max}$ , can be determined quantitatively from the  $P(r)$  function as  $P(r)$  drops to zero at  $r = d_{max}$ . The radius of gyration,  $R_g$ , was estimated from the second moment of the  $P(r)$  function by the equation:

$$R_g^2 = \frac{\int_0^{d_{max}} r^2 P(r) dr}{2 \int_0^{d_{max}} P(r) dr} \quad (7)$$

A series of concentrations were measured for each protein at sufficiently dilute concentrations (1 to 5 mg/ml). The measured values ( $R_g$ ,  $d_{max}$ , and  $I(0)/c$ ) did not differ significantly or in a systematic manner within each series, indicating that intermolecular interactions were negligible at these concentrations. The value reported for each protein is an average of the values determined at various concentrations.

### Acknowledgements

This work was supported in part by grants from the National Institutes of Health to J.E.C. (DK09070), to

D.M.E. (GM22778), and to D.G.S. (AI32524). D.G.S. is an Assistant Investigator of the Howard Hughes Medical Institute. National Institutes of Health NRSA postdoctoral fellowships are acknowledged by K.K.R. (GM16039) and K.G.F. (GM16769).

## References

- Bach, I., Rhodes, S. J., Pearse, R. V., Heinzl, T., Gloss, B., Scully, K. M., Sawchenko, P. E. & Rosenfeld, M. G. (1995). P-Lim, a LIM homeodomain F factor, is expressed during pituitary organ and cell commitment and synergizes with Pit-1. *Proc. Natl Acad. Sci. USA*, **92**, 2720–2724.
- Barlow, P. N., Luisi, B., Milner, A., Elliot, M. & Everett, R. (1994). Structure of the C<sub>3</sub>HC<sub>4</sub> domain by <sup>1</sup>H-nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **237**, 201–211.
- Berg, J. M. (1990). Zinc fingers and other metal-binding domains. *J. Biol. Chem.* **265**, 6513–6516.
- Borden, K. L. B., Boddy, M. N., Lally, J., O'Reilly, N. J., Martin, S., Howe, K., Solomon, E. & Freemont, P. S. (1995). The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J.* **14**, 1532–1541.
- Cheng, G., Cleary, A. M., Ye, Z.-s., Hong, D. I., Lederman, S. & Baltimore, D. (1995). Involvement of CRAF1, a relative of TRAF, in CD40 signaling. *Science*, **267**, 1494–1498.
- Cohn, E. J. & Edsall, J. T. (1943). Density and apparent specific volume of proteins. In *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions* (Cohn, E. J. & Edsall, J. T., eds), pp. 370–381, Reinhold Publishing Corporation, New York.
- Coleman, J. E. (1992). Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins. *Annu. Rev. Biochem.* **61**, 897–946.
- Cortes, P., Ye, Z. S. & Baltimore, D. (1994). RAG-1 interacts with the repeated amino acid motif of the human homologue of the yeast protein SRP1. *Proc. Natl Acad. Sci. USA*, **91**, 7633–7637.
- Cuomo, C. A. & Oettinger, M. A. (1994). Analysis of regions of RAG-2 important for V(D)J recombination. *Nucl. Acids Res.* **22**, 1810–1814.
- Engelman, D. M. & Moore, P. B. (1975). Determination of quaternary structure by small angle neutron scattering. *Annu. Rev. Biophys. Bioeng.* **4**, 219–241.
- Feuerstein, R., Wang, X., Song, D., Cooke, N. E. & Liebhaber, S. A. (1994). The LIM/double zinc-finger motif functions as a protein dimerization domain. *Proc. Natl Acad. Sci. USA*, **91**, 10655–10659.
- Frankel, A. D., Berg, J. M. & Pabo, C. O. (1987). Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc. Natl Acad. Sci. USA*, **84**, 4841–4845.
- Freemont, P. S. (1993). The RING finger: a novel protein sequence motif related to the zinc finger. *Ann. N.Y. Acad. Sci.* **684**, 174–192.
- Freemont, P. S., Hanson, I. M. & Trowsdale, J. (1991). A novel cysteine-rich sequence motif. *Cell*, **64**, 483–484.
- Futreal, P. A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S. *et al.* (1994). *BRCA1* mutations in primary breast and ovarian carcinomas. *Science*, **266**, 120–122.
- Gardner, K. H., Anderson, S. F. & Coleman, J. E. (1995). Solution structure of the *Kluyveromyces lactis* LAC9 Cd<sub>2</sub>Cys<sub>6</sub> DNA-binding domain. *Nature Struct. Biol.* **2**, 898–905.
- Glatter, O. (1982). In *Small Angle X-ray Scattering* (Glatter, O. & Kratky, O., eds), pp. 119–165, Academic Press, London.
- Gorlich, D., Prehn, S., Laskey, R. A. & Hartmann, E. (1994). Isolation of a protein that is essential for the first step of nuclear protein import. *Cell*, **79**, 767–778.
- Hsieh, C. & Lieber, M. R. (1992). CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication. *EMBO J.* **11**, 315–325.
- Jasanoff, A., Kochoyan, M., Fraenkel, E., Lee, J. P. & Weiss, M. A. (1992). Aromatic-aromatic interactions in the zinc finger motif. Analysis of the two-dimensional nuclear magnetic resonance structure of a mutant domain. *J. Mol. Biol.* **225**, 1035–1047.
- Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* **36**, 575–588.
- Kuntz, I. D. (1971). Hydration of macromolecules III. Hydration of polypeptides. *J. Am. Chem. Soc.* **93**, 514–516.
- Laue, T. M., Shah, B., Ridgeway, T. M. & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds), pp. 90–125, Royal Society of Chemistry, Cambridge, UK.
- Leu, T. M. J. & Schatz, D. G. (1995). Rag-1 and Rag-2 are components of a high molecular weight complex and association of Rag-2 with this complex is Rag-1 dependent. *Mol. Cell. Biol.* **15**, 5657–5670.
- Lewis, S. M. (1994). The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative analyses. *Advan. Immunol.* **56**, 27–150.
- Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R. & Sigler, P. B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature*, **352**, 497–505.
- Maina, C. V., Riggs, P. D., Granda, A. G., Slatko, B. E., Moran, L. S., Tagliamonte, J. A., Reynolds, L. A. & Guan, C. D. (1988). An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene*, **74**, 365–373.
- McBlane, J. F., van Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M. & Oettinger, M. A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell*, **83**, 387–395.
- McRorie, D. K. & Voelker, P. J. (1993). *Self-associating Systems in the Analytical Ultracentrifuge*. Beckman Instruments, Inc., Fullerton, CA.
- Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science*, **248**, 1517–1523.
- Omichinski, J. G., Clore, G. M., Robien, M., Sakaguchi, K., Apella, E. & Gronenborn, A. M. (1992). High resolution solution structure of the double Cys<sub>2</sub>His<sub>2</sub> zinc finger from the human enhancer binding protein MBP-1. *Biochemistry*, **31**, 3907–3917.
- Pan, T. & Coleman, J. E. (1990). The DNA binding domain of GAL4 forms a binuclear metal ion complex. *Biochemistry*, **29**, 3023–3029.
- Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T. & Klevit, R. E. (1988).

- Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science*, **241**, 1489–1492.
- Pavletich, N. P. & Pabo, C. O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science*, **261**, 1701–1707.
- Perez-Alvarado, G. C., Miles, C., Michelson, J. W., Louis, H. A., Winge, D. R., Beckerle, M. C. & Summers, M. F. (1994). Structure of the C-terminal LIM domain from the cysteine rich protein CRP. *Nature Struct. Biol.* **1**, 388–398.
- Qian, X. & Weiss, M. A. (1992). Two-dimensional NMR studies of the zinc finger motif: solution structures and dynamics of mutant ZFY domains containing aromatic substitutions in the hydrophobic core. *Biochemistry*, **31**, 7463–7476.
- Rodgers, K. K. & Coleman, J. E. (1994). DNA binding and bending by the transcription factors GAL4(62\*) and GAL4(149\*). *Protein Sci.* **3**, 608–619.
- Rothe, M., Sarma, V., Dixit, V. M. & Goeddel, D. V. (1995). TRAF2-mediated activation of NF- $\kappa$ B by TNF receptor 2 and CD40. *Science*, **269**, 1424–1427.
- Sadofsky, M. J., Hesse, J. E., McBlane, J. F. & Gellert, M. (1993). Expression and V(D)J recombination activity of mutated RAG-1 proteins. *Nucl. Acids Res.* **21**, 5644–5650.
- Sadofsky, M. J., Hesse, J. E. & Gellert, M. (1994). Definition of a core region of RAG-2 that is functional in V(D)J recombination. *Nucl. Acids Res.* **22**, 1805–1809.
- Schatz, D. G. & Leu, T. M. J. (1996). Rag-1 and rag-2: biochemistry and protein interactions. In *Molecular Analysis of DNA Rearrangements in the Immune System* (Jessberger, R. & Lieber, M. R., eds), Springer-Verlag, Heidelberg. In the press.
- Schatz, D. G., Oettinger, M. A. & Baltimore, D. (1989). The V(D)J recombination activating gene, *RAG-1*. *Cell*, **59**, 1035–1048.
- Schmeichel, K. L. & Beckerle, M. C. (1994). The LIM domain is a modular protein-binding interface. *Cell*, **79**, 211–219.
- Schwabe, J. W. R. & Klug, A. (1994). Zinc mining for protein domains. *Nature Struct. Biol.* **1**, 345–349.
- Silver, D. P., Spanopoulou, E., Mulligan, R. C. & Baltimore, D. (1993). Dispensable sequence motifs in the *RAG-1* and *RAG-2* genes for plasmid V(D)J recombination. *Proc. Natl Acad. Sci. USA*, **90**, 6100–6104.
- Semenyuk, A. V. & Svergun, D. I. (1991). GNOM—a program package from small-angle scattering data processing. *J. Appl. Crystallog.* **24**, 537–540.
- Spanopoulou, E., Cortes, P., Shih, C., Huang, E., Silver, D. P., Svec, P. & Baltimore, D. (1995). Localization, interaction and RNA binding properties of the V(D)J recombination-activating proteins Rag-1 and Rag-2. *Immunity*, **3**, 715–726.
- Spurlino, J. C., Lu, G.-Y. & Quijcho, F. A. (1991). The 2.3-Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.* **266**, 5202–5219.
- Sreerama, N. & Woody, R. W. (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **209**, 32–44.
- Stafford, W. (1992). Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile. *Anal. Biochem.* **203**, 295–391.
- van Gent, D. C., McBlane, J. F., Ramsden, D. A., Sadofsky, M. J., Hesse, J. E. & Gellert, M. (1995). Initiation of V(D)J recombination in a cell-free system. *Cell*, **81**, 925–934.

*Edited by P. E. Wright*

*(Received 16 January 1996; received in revised form 5 April 1996; accepted 18 April 1996)*