THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 23, Issue of June 6, pp. 21232–21239, 2003

Printed in U.S.A.

Structural Requirements for Assembly of the CSL·Intracellular Notch1·Mastermind-like 1 Transcriptional Activation Complex*

Received for publication, February 13, 2003
Published, JBC Papers in Press, March 18, 2003, DOI 10.1074/jbc.M301567200

Yunsun Nam, Andrew P. Weng, Jon C. Aster, and Stephen C. Blacklow‡

From the Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115

Ligand binding by Notch receptors triggers a series of proteolytic cleavages that liberate the intracellular portion of Notch (ICN) from the cell membrane, permitting it to translocate to the nucleus. Nuclear ICN binds to a highly conserved DNA-binding transcription factor called CSL (also known as RBP-Jκ, CBF1, Suppressor of Hairless, and Lag-1) and recruits Mastermind-like transcriptional co-activators to form a transcriptional activation complex. Using bioinformatics tools, we identified a Rel homology region (RHR) within CSL that was used as a guide to determine the minimal protein requirements for ternary complex formation. The RHR of CSL contains both the N- and C-terminal β-sheet domains (RHR-n and RHR-c) of typical Rel transcription factors, as judged by circular dichroism spectra. Binding of monomeric CSL to DNA requires the entire RHR of CSL and an additional 125-residue N-terminal sequence, whereas binding to ICN requires only the RHR-n domain. Although the RAM (RBP-Jκ (recombination-signal-sequence-binding protein for Jκ genes)-associated molecule) domain of ICN is flexible and relatively unstructured as an isolated polypeptide in solution, it associates stably with CSL on DNA. Recruitment of Mastermind-like 1 (MAML1) to CSL·ICN complexes on DNA requires inclusion of the ankyrin repeat domain of ICN, and N- and C-terminal sequences of CSL extending beyond the DNA-binding region. The requirement for cooperative assembly of the MAML1·ICN·CSL·DNA complex suggests that a primary function of ICN is to render CSL competent for MAML loading. On the basis of our results, we present a working structural model for the organization of the MAML1·ICN·CSL·DNA complex.

Notch proteins act as receptors in an evolutionarily conserved signaling pathway that controls differentiation and proliferation in response to ligands expressed on neighboring cells (reviewed in Refs. 1 and 2). Ligand binding to the extracellular region of Notch receptors causes activation of a signal by triggering a series of proteolytic cleavages that release the intracellular portion of Notch (ICN)[1] from the plasma membrane. ICN then migrates to the nucleus where it activates the transcription of target genes.

ICNs are modular polypeptides comprised of an N-terminal RAM domain of ~110 amino acids, a set of seven iterated ankyrin repeats (3, 4), and less conserved C-terminal regions that include a transcriptional activation domain and a far C-terminal PEST sequence. The ankyrin repeat (ANK) domain is the most highly conserved region of ICN and is essential for all known Notch functions (1).

The primary target of ICN in the nucleus is the highly conserved DNA-binding transcription factor, CSL (also known as RBP-Jκ, CBF1, Suppressor of Hairless (Su)H), and Lag-1) (5–7). Previous analysis of CSL proteins using primary sequence alignment tools has failed to detect similarity to other known transcription factors. In the absence of ICN, CSL represses transcription by virtue of interactions with a number of co-repressors, including SMRT/N-CoR (silencing mediator for retinoic acid and thyroid hormone receptors/nuclear repressor co-repressor), CIR (CBF1-interacting co-repressor), and KyoT2 (8–10).

Binding of ICNs not only displaces co-repressors from CSL proteins but also recruits transcriptional co-activators to CSL·ICN complexes on DNA. Several studies indicate that Mastermind-like polypeptides such as Lag-3 in worms (11) and mammalian Mastermind-like-1 (MAML1) are core components of a higher order transcriptional activation complex on DNA (12). The existence of a CSL·ICN·Mastermind signaling complex was foreshadowed by genetic screens in the fly revealing that mutations in each of these genes cause similar "neurogenic" loss-of-function phenotypes (13–15).

Although the level of ICN in the nucleus of most cells undergoing physiologic Notch signaling is too low to be detected readily, certain T cell leukemias associated with chromosomal translocations involving the human *NOTCH1* gene express nuclear ICN1-like polypeptides at high levels (16). Capobianco's group (17) exploited a cell line derived from a *NOTCH1* leukemia, SUPT-1, to demonstrate the presence of endogenous high molecular mass (>1 MDa) nuclear complexes containing ICN1, MAML1, and CSL. We have also shown that the effect of disruption of MAML1 recruitment on Notch-dependent cells resembles that of inhibition of γ-secretase cleavage of membrane-tethered Notch (18). These studies and previous genetic and functional data suggest that CSL·ICN·MAML1 is an essential subcomplex within a larger protein assembly required for the transcription of Notch target genes.

* This work was supported by National Institutes of Health Grants HL61001 and CA92433 (to S. C. B.) and CA82308 (to J. C. A.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Pew Scholar in the biomedical sciences and an Established Investigator of the American Heart Association. To whom correspondence should be addressed. Tel.: 617-732-5799; Fax: 617-264-5296; E-mail: sblacklow@rics.bwh.harvard.edu.

[1] The abbreviations used are: ICN, intracellular portion of Notch; ANK, ankyrin; RAM, RBP-J-associating molecule; MAML, mammalian Mastermind-like; RHR, Rel homology region; RHR-n and -c, RHR N- and C-terminal β-sheet domains; IVT, *in vitro* transcription/translation; GST, glutathione *S*-transferase; TRX, thioredoxin; TEV, tobacco etch virus; DTT, dithiothreitol; Ni-NTA, nickel-nitrilotriacetic acid; EMSA, electrophoretic mobility shift assay; 3D-PSSM, three-dimensional position-specific scoring matrix; NFAT, nuclear factor of activated T cells.

MAML1 and other members of the MAML family are polypeptides of ~1000 amino acid residues predicted to be of low structural complexity, suggesting that they function as scaffolds for further recruitment of additional co-activators and/or the transcription machinery. The N-terminal region of MAML1 and related polypeptides contains a sequence that associates with ICN and CSL (12, 19). When expressed as a truncated moiety, the N-terminal ICN-interaction region of MAML1 is a potent dominant negative inhibitor of Notch signals, indicating that C-terminal sequences are also needed for proper function (12). Recent studies indicate these C-terminal sequences bind the transcriptional co-activator p300 as well as other unknown factors that are needed for activation of transcription from chromatinized templates *in vitro* (20, 21).

At present there is limited information about the structure of CSL and its higher order complexes with ICN and MAML1. Here, using bioinformatics tools, we have located the two canonical domains of a previously unidentified Rel homology region (RHR) within CSL, suggesting that CSL is a distant relative of the Rel homology family of transcription factors. With these domains as a guide, we define the regions of CSL required for the formation of subcomplexes and for assembly of ternary complexes on DNA, showing that both ICN and CSL must be present to recruit MAMLl. These mapping studies support the classification of CSL as a member of the Rel homology family of transcription factors. The requirement for cooperative assembly of the ternary complex on DNA, combined with the CSL mapping studies, suggests that a primary function of ICN1 is to render CSL competent for MAML1 loading.

EXPERIMENTAL PROCEDURES

DNA Constructs—All CSL constructs are derived from isoform 3 of human CSL. CSL cDNAs used to produce polypeptides using *in vitro* transcription/translation (IVT) reactions were amplified by PCR and purified by gel extraction. IVT reactions were performed using TNT® T7 Quick for PCR-DNA (Promega) according to the manufacturer's instructions. For bacterial expression, cDNAs encoding GST-CSL RHR (residues Lys-156–Gly-435), GST-CSL RHR-n (residues Lys-156–Pro-332), and GST-CSL RHR-c (residues Ala-335–Gly-435) were assembled in the vector pGEX-4T-1. Each construct contained a tobacco etch virus (TEV) protease recognition site (ENLYFQG) at the junction between GST and the CSL sequence. cDNAs encoding the polypeptides Arg-1761—Gly-2127 (RAMANK) and Met-1873—Gly-2127 (ANK) of human Notch1 were inserted in frame downstream of GST and TEV cleavage site coding sequences in a modified version of the vector pDEST15 using GATEWAY technology (Invitrogen). A cDNA encoding the polypeptide Arg-1761—Ser-1890 (RAM) was inserted in-frame between a similar 5′-GST-TEV site coding sequence and a 3′-GTHHHHHH tag in a modified pET vector. A cDNA encoding residues Leu-13–His-74 of human MAML1 (MAML1-(13–74)) was inserted in-frame downstream of a hexahistidine tag and TEV cleavage site. Expression constructs for full-sized MAML1 (12), ICN1, and CSL (22) have been described previously.

Protein Purification—GST-CSL fusion protein expression was induced for 3 h in *Escherichia coli* strain BL21(DE3) pLysS by the addition of isopropyl-1-thio-β-D-galactopyranoside (0.5 mM) during log phase. Cells were collected by centrifugation and lysed by sonication in 50 mM Tris, pH 8.0, containing 150 mM NaCl, sucrose (20% w/v), 5 mM dithiothreitol (DTT), 1 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 2 μg/ml aprotinin, and 0.7 μg/ml pepstatin. Solubilized proteins were adsorbed to glutathione-Sepharose beads, which were washed with buffer A (50 mM Tris, pH 8.0, 150 mM NaCl, and 5 mM β-mercaptoethanol) and incubated with 500 units of recombinant His-tagged TEV protease (Invitrogen) at 25 °C for 36–48 h. Released RHR-c was further purified by anion exchange chromatography on Mono-Q resin (Amersham Biosciences) using a linear gradient of NaCl (0–1 M) in 20 mM Tris buffer, pH 8.0, containing 5 mM DTT and 1 mM EDTA. In a final step, RHR and RHR-c were purified to >95% homogeneity by size exclusion chromatography on a Superdex 200 column (Amersham Biosciences) in 50 mM Tris buffer, pH 8.0, containing 150 mM NaCl, 5 mM DTT, and 1 mM EDTA. Notch1 polypeptides were expressed and purified similarly. After cleavage, His-tagged TEV was cleared from soluble RAMANK and ANK by adsorption to Ni-NTA-agarose beads (Qiagen). Soluble RAM-H_6 was bound to Ni-NTA-agarose beads and eluted in buffer A containing 250 mM imidazole. All three proteins were then further purified by anion exchange chromatography on Mono-Q resin (Amersham Biosciences) using a linear gradient of NaCl (0–1 M) in 20 mM Tris buffer, pH 8.0, containing 5 mM DTT and 1 mM EDTA. Finally, ANK and RAMANK were further purified by size exclusion chromatography on a Superdex 200 column (Amersham Biosciences) in 50 mM Tris buffer, pH 8.0, containing 150 mM NaCl, 5 mM DTT, and 1 mM EDTA.

Expression of H_6-tagged MAML1-(13–74) was induced as described above for CSL and ICN polypeptides. H_6-MAML-(13–74) was purified from inclusion bodies by solubilization in 50 mM Tris buffer, pH 8.0, containing 6 M guanidinium HCl and 150 mM NaCl, followed by affinity chromatography on a Ni-NTA-Sepharose column. After extensive dialysis against 5% acetic acid, the protein was lyophilized, dissolved in 50 mM sodium citrate buffer, pH 4.9, containing 5 mM DTT and 0.5 mM EDTA, and cleaved with TEV protease. Free MAML1-(13–74) polypeptide was purified further by cation exchange chromatography (MonoS 5/5 HR, Amersham Biosciences) using a gradient of NaCl (0–1 M) in 20 mM sodium phosphate buffer, pH 6.8, containing 5 mM DTT and 1 mM EDTA. Purified MAML1-(13–74) was dialyzed against 5% acetic acid and lyophilized. Before use, MAML1-(13–74) was dissolved in 50 mM Tris HCl, pH 8.0, buffer with 150 mM NaCl and 5 mM DTT.

Bioinformatics—Multiple sequence alignments were performed using CLUSTAL-W.[2] Secondary structure prediction for all protein sequences was carried out using the PHD program (23, 24). The structural homology of CSL proteins (human CSL and *Drosophila* Suppressor of Hairless) to Rel proteins of known three-dimensional structure was detected using the 3D-PSSM fold-recognition program[3] (25, 26).

Biophysical Studies—Circular dichroism (CD) spectra were acquired on an Aviv 62DS spectropolarimeter equipped with a thermoelectric temperature controller. Protein concentrations were determined according to Edelhoch (27). Samples of CSL RHR-c (36 μM), RAM (10 μM), ANK (10 μM), and RAMANK (10 μM) were prepared in 10 mM sodium phosphate buffer, pH 7.4, containing 0.2 mM DTT and 150 mM NaCl. CSL RHR (24 μM) was prepared in the same buffer with 300 mM NaCl. CD spectra of CSL RHR, RAM, ANK, and RAMANK were recorded at 4 °C in a 1-mm path length cuvette as the average of five scans using a 3-s acquisition time. The CSL RHR-c sample was scanned 10 times with an acquisition time of 30 s. Thermal melts of each Notch1 polypeptide (5 μM) were performed in a 1-cm path length cuvette in the same buffer. The CD signal was monitored at 222 nm using a 2-min equilibration period, a 30-s signal acquisition time, and a step size of 1 °C from 4 to 75 °C. The thermal denaturation curves of RAMANK and ANK were ~95% reversible. The T_m for unfolding was calculated by finding the minimum of the first derivative of the CD signal with respect to $1/T$.

Equilibrium sedimentation measurements were performed on a Beckman Optima XL-A analytical Ultracentrifuge using an An-Ti rotor and six-sector equilibrium centerpieces. To determine the masses of ANK and RAMANK, rotor speeds of 10,000, 15,000, and 20,000 rpm were used. Samples of ANK and RAMANK, prepared in 10 mM sodium phosphate buffer, pH 7.4, containing 150 mM NaCl and 0.2 mM DTT, were dialyzed exhaustively against the same buffer. Protein samples (ANK: 1.8 and 6 μM; RAMANK: 1 and 3.3 μM) were prepared by dilution with dialysate, which was also used to fill the reference channels. Absorbance data were acquired at three wavelengths (229, 236, and 280 nm) by averaging 10 scans with a 0.001-cm radial step size. Molecular weights were calculated by fitting data sets to a single ideal species model using Microsoft Excel software and partial specific volumes of 0.726 for ANK and 0.724 for RAMANK (28). No systematic residuals were observed.

Electrophoretic Mobility Shift Assay—Electrophoretic mobility shift assays (EMSAs) were performed using a ^{32}P-labeled CSL-specific oligonucleotide probe as described previously (12).

GST Bead Binding Assays—Glutathione-Sepharose beads were mixed at 4 °C for 1 h with cleared bacterial lysates containing GST fusion polypeptides. After five washes with 50 mM Tris-HCl buffer, pH 8.0, containing 150 mM NaCl, 5 mM DTT, and 0.5% Nonidet P-40 (v/v), beads with GST fusion proteins (20–30 μl/experiment) were incubated with potential binding partners for 1 h at 4 °C followed by five washes in the same buffer. Bound proteins were released by heating (100 °C) for 10 min in SDS-PAGE loading buffer. After electrophoresis in SDS-

[2] www.ebi.ac.uk/clustalw.

[3] www.sbg.bio.ic.ac.uk/~3dpssm/.

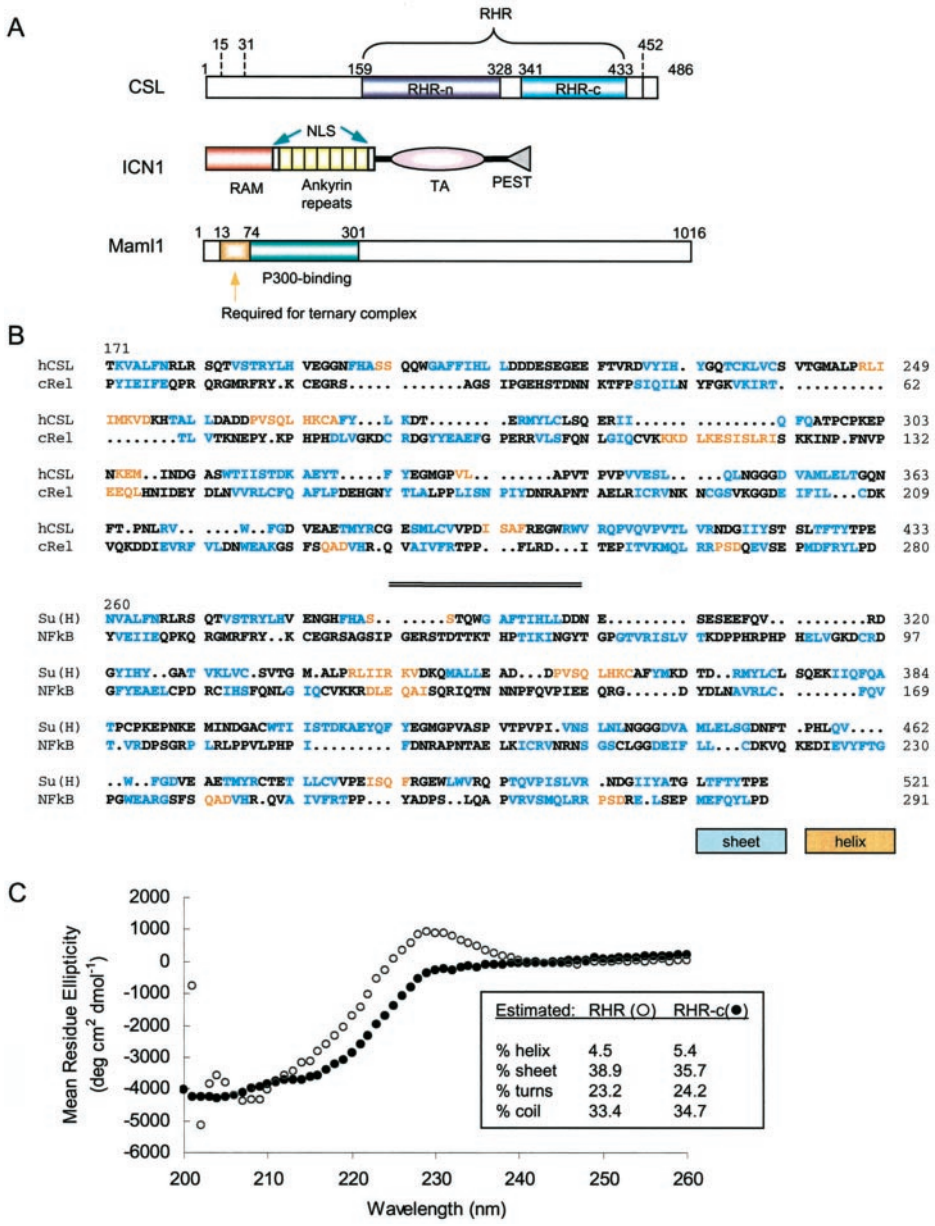

FIG. 1. *A*, domain organization of CSL, ICN1, and Maml1. *CSL*: the homology of human CSL with *Drosophila* Su(H) starts at residue 15 and with RBP-L at residue 31. The RHR-n and RHR-c domains predicted by 3D-PSSM are shown. *ICN1*: *NLS*, nuclear localization sequence (*arrows*); *TA*, transactivation region. *Maml1*: residues 13–74 correspond to the minimum length polypeptide necessary and sufficient to form ternary complexes with ICN, CSL, and DNA. Residues 75–301 of MAML1 are implicated in the recruitment of CBP (CREB-binding protein)/p300 (20). *B*, alignment of the RHR regions of CSL and Su(H) with their structural homologues. 3D-PSSM alignments of CSL with c-*rel* (*top*) and Su(H) with NF-κB (*bottom*). Residues are colored by the predicted (CSL and Su(H)) or actual (c-*rel* and NF-κB) secondary structure (*blue*, β-strands; *orange*, α-helices). *C*, CD spectra of the entire RHR of CSL and of CSL RHR-c. *Open circles*, entire RHR; *closed circles*, RHR-c. The estimated of protein secondary structure content was calculated with the program CONTINLL (43).

polyacrylamide gels, proteins were visualized by Coomassie Blue staining.

RESULTS

Detection of the Rel Homology Regions of CSL—The major isoform of human CSL encodes a protein of 486 amino acids (Fig. 1*A*). Known biochemical activities of CSL include sequence-specific DNA binding (29), association with ICN polypeptides (5), and formation of higher order complexes with MAML polypeptides (11, 12). To investigate the structural basis for these activities, we used a bioinformatics approach to guide the design of CSL constructs. Because significant primary sequence homology between CSL proteins and other proteins in the data base was not detectable, we searched for evidence of homology to protein domains of known structure using the 3D-PSSM fold-recognition program, which compares input sequences with the existing protein fold library using structure-based sequence profiles and secondary structure prediction (25, 26). 3D-PSSM returned the RHR of c-*rel* as the top hit for human CSL (residues 171–433; *E* value, 0.121 (where *E* is theoretical expectation value)). Moreover, the *Drosophila* orthologue of CSL, SU(H), also retrieved an RHR protein, the p65 subunit of NF-κB, with a confidence level of more than 90%. Both query sequences retrieved other RHR-containing proteins as well, such as TonEBP (residues 159–433 of CSL; *E* value, 0.193) and NFAT1. These results suggest that CSL proteins are distantly related to the Rel homology family of transcription factors (Fig. 1, *A* and *B*).

The RHR, a conserved region among proteins of the Rel family (30), consists of two immunoglobulin-like folds of β-sheets (31, 32) termed RHR-n and RHR-c (also known as a transcription factor Ig-like (TIG) domain), predicted by 3D-PSSM to correspond to residues 159–328 and 341–433, respectively, of CSL. The RHR-c folds are also identified in both CSL proteins by PFAM (protein families data base of alignments and hidden Markov models), but the RHR-n folds are not. Because the structures of previously characterized RHR-n domains are more highly divergent than the RHR-c domains, a poorer alignment of the CSL RHR-n domain is not surprising or unexpected. Moreover, the convincing prediction of the RHR-c domain itself argues strongly for the coexistence of an accompanying RHR-n domain.

To test directly whether the predicted RHR region of CSL

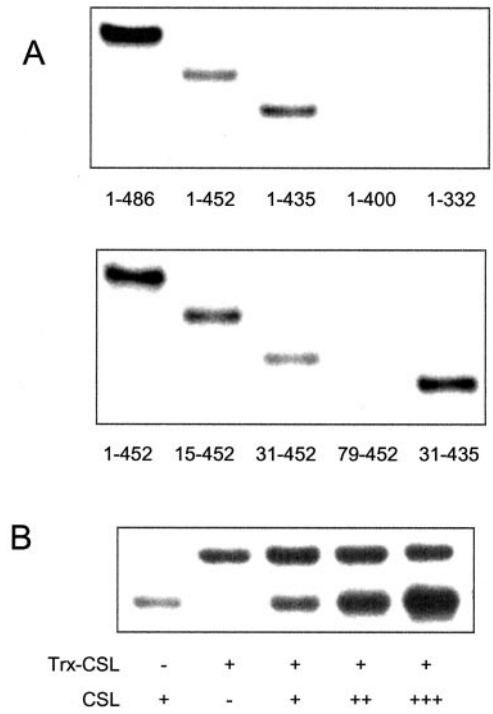

FIG. 2. **CSL requires residues 31–435 to bind DNA as a monomer.** *A*, CSL variants corresponding to the indicated amino acid residues were tested for binding to a ^{32}P-labeled oligonucleotide containing a consensus CSL binding site in an EMSA. DNA·protein complexes were detected by autoradiography. *B*, CSL-(1–452) (*CSL*) and CSL-TRX (*Trx-CSL*), polypeptides of differing size, were tested by EMSA, alone and in combination, as described in *A*.

contains two separate domains in which the predominant structural elements are β-sheets, we estimated the secondary structure content of the predicted RHR and RHR-c domains by CD. These measurements show that the predicted RHR region of CSL, as well as the isolated RHR-c domain, contains a predominance of β-stranded secondary structure and little α-helix (Fig. 1*C*). With the 3D-PSSM predictions, the CD findings are consistent with the presence of a complete RHR domain within CSL proteins.

The predicted RHR domain boundaries were combined with additional information to guide subsequent structure-function studies. Truncation points outside of the RHR domain boundaries were chosen on the basis of primary sequence homology and secondary structure prediction. Homology between CSL and RBP-L, a related protein that binds the same DNA sequence but not ICNs (33), starts at residue 31. C-terminal to the RHR of CSL, residue 452 marks the end of a stretch of 20 residues (433–452) predicted by PHD (23, 24) to contain helical structure. The C-terminal 34 residues 453–486 consist of a low complexity sequence with little predicted secondary structure. Finally, sites chosen for intradomain truncations (*e.g.* residues 79 and 400) lie within regions predicted by PHD to lack secondary structure. Guided by the informatics results, we therefore made CSL expression constructs with N termini starting at positions 31, 79, and 156 and C-terminal truncations at positions 332, 400, 435, and 452.

CSL Requires Residues 31–435 to Bind DNA as a Monomer—Using these constructs, we first investigated the requirements for binding of CSL to DNA using an EMSA (Fig. 2*A*). Removal of the first 31 residues of CSL does not interfere with DNA binding, but constructs with N termini beginning at residue 79 or 156 fail to bind DNA (Fig. 3*A*), indicating that the RHR region alone is insufficient. C-terminal truncation of CSL to residue 452 or 435 does not interfere with DNA binding, but further truncations to residue 400 or 332, which interrupt or delete the RHR-c domain, abrogate DNA binding. Thus, CSL residues 31–435, which encompass the complete RHR and an additional N-terminal extension of 125 amino acids, are both necessary and sufficient to bind DNA.

Because other Rel homology transcription factors related to NF-κB bind to DNA as dimers (31, 32, 34), we revisited the question of stoichiometry in CSL·DNA complexes by performing EMSA in the presence of a mixture of two CSL protein variants of different lengths: CSL-(1–452) and CSL-(1–452) bearing a C-terminal thioredoxin tag (CSL-TRX). Each form of CSL alone binds DNA, and the two CSL·DNA complexes migrate differently in the EMSA (Fig. 2*B*). When both CSL species are bound to DNA, heterodimers (which would give an electrophoretic mobility shift between CSL and CSL-TRX) are not observed, confirming that CSL binds DNA as a monomer, as reported previously for mouse CSL using nuclear extracts (35).

Biophysical Characterization of RAM, ANK, and RAMANK—Sequence alignment, secondary structure prediction tools, and comparison with BCL-3 (36), the closest homologue of the ANK domain of ICN1 in the structure data base, suggest that residues 1873–2127 of human Notch1 comprise a series of seven ankyrin repeats (Fig. 1*A*), a prediction supported by biophysical studies of the ANK domain of *Drosophila* Notch (3, 4). The RAM region, however, has no detectable sequence homology to any protein of known structure. Therefore, we prepared highly purified recombinant RAM, ANK, and RAMANK for biophysical studies to evaluate the integrity and stability of the ankyrin repeat domain of human Notch1 and to determine whether the RAM region contains a stable ordered structure in isolation or in its native context (Fig. 3*A*).

As judged by circular dichroism, the ANK polypeptide contains about 40% α-helix, an amount within the range predicted for an ankyrin repeat domain (Fig. 3*B*). ANK undergoes a cooperative, reversible thermal unfolding transition, with a T_m of 50 °C (Fig. 3*C*). Equilibrium sedimentation indicates that ANK does not exhibit self-associating behavior at concentrations of up to at least 6 μM and is monomeric (measured molecular mass = 33.5 kDa, calculated molecular mass = 27.9 kDa; Fig. 3*D*) at the concentrations used in the thermal unfolding experiments.

In contrast, the RAM region is predominantly unstructured, as judged by CD (Fig. 3*B*). Furthermore, inclusion of the RAM region in the context of RAMANK does not induce a significant amount of α or β structure in the RAM region, because the calculated difference spectrum, which represents the signal from the RAM region in the context of RAMANK, closely approximates the spectrum of RAM-H_6 alone (Fig. 3*B*). The T_m for unfolding of RAMANK is the same as for unfolding of ANK, indicating that the presence of RAM does not detectably influence the stability of ANK (Fig. 3*C*). Thus, the RAM region, both as an isolated polypeptide and in the context of the adjacent ANK domain, remains a flexible polypeptide segment without significant secondary structure. RAMANK is monomeric and does not exhibit self-associating behavior at concentrations up to at least 3.3 μM, as judged by equilibrium sedimentation (measured molecular mass = 46.0 kDa; calculated molecular mass = 40.7; Fig. 3*E*).

The RHR-n Region of CSL Forms Binary Complexes with ICN—Next, we determined the region of CSL required for binding of ICN. Previous mammalian two-hybrid assays showed residues 156–332 of CSL interacting with ICN (37), which suggests that this interaction is mediated through the predicted RHR-n domain (159–328). Indeed, CSL-(156–332) expressed in IVT reactions associates with GST-RAMANK

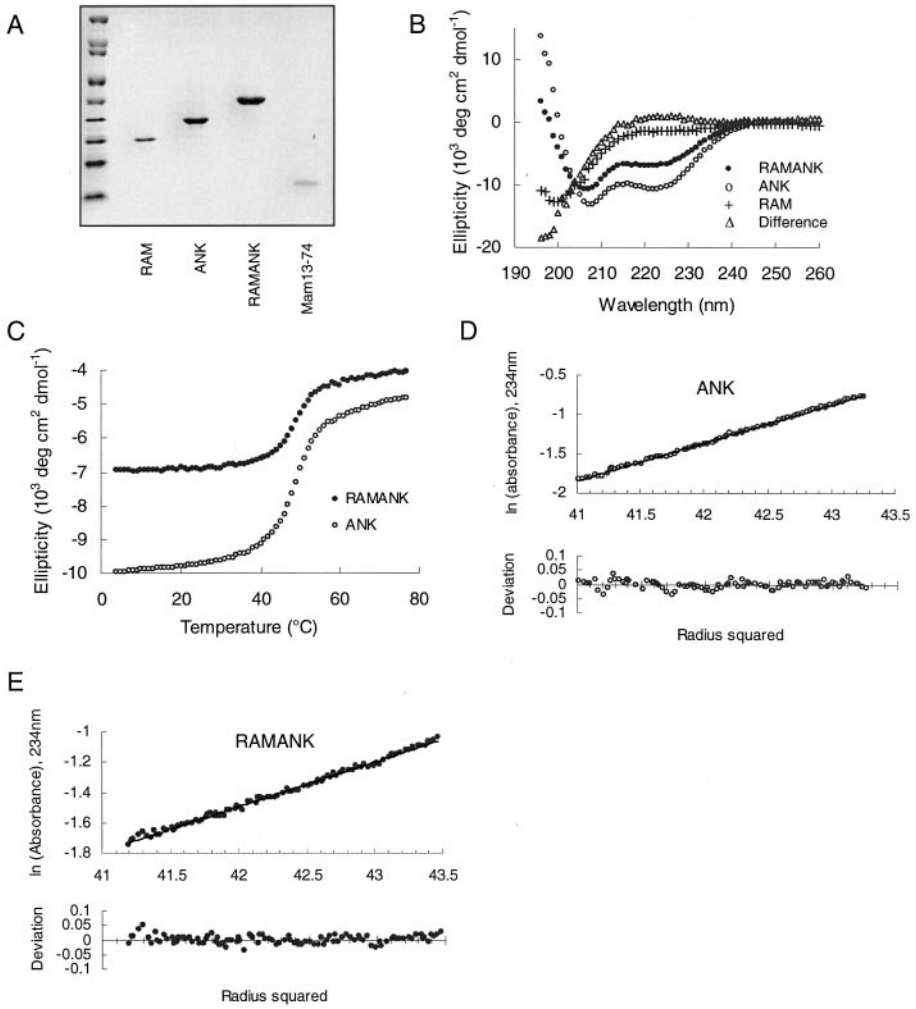

FIG. 3. **Purification and characterization of RAMANK, RAM, ANK, and MAML1-(13–74).** *A*, Coomassie staining of purified RAM, ANK, RAMANK, and MAML-13–74 after SDS-PAGE. *B*, CD spectra of RAM (*crosses*), ANK (*open circles*), and RAMANK (*closed circles*). Spectra were recorded at 4 °C in physiologic buffer (10 mM sodium phosphate, pH 7.4, 150 mM NaCl, 0.2 mM DTT). A predicted CD spectrum of the RAM region (*open triangles*) was calculated from the difference between the RAMANK and ANK spectra. *C*, thermal dependence of the CD signal at 222 nm for ANK (*open circles*) and RAMANK (*closed circles*) in physiologic buffer. *D* and *E,* equilibrium centrifugation of 6 μM ANK (*D*) and 3.3 μM RAMANK (*E*) at 4 °C in physiologic buffer (10 mM sodium phosphate, pH 7.4, 150 mM NaCl, 0.2 mM DTT). *Upper panels*, log of the concentration distribution of protein as a function of the square of the radial position at 15,000 rpm (*ANK*) or 10,000 rpm (*RAMANK*). The best-fit line describes a single homogeneous species of molecular mass 33,818 Da (*ANK*) or 46,008 Da (*RAMANK*). *Lower panels*, residuals to the fit. No systematic trend is observed.

(Fig. 4*A*). To verify that this binding is direct, we purified recombinant GST-CSL-(156–332) from bacteria and performed pull-down assays on glutathione beads with recombinant purified RAMANK, RAM, and ANK. Recombinant GST-CSL-(156–332) pulls down both RAMANK and RAM, indicating that these binary complexes are stable in the absence of additional protein factors or DNA (Fig. 4*B*). In contrast, ANK does not form a stable complex with CSL-(156–332), either because the affinity of the isolated RHR-n domain for ANK is too weak to detect in the absence of other factors (*i.e.* MAML1 and/or DNA) or because ANK only contacts RHR-c.

Requirements for CSL/ICN/MAML1 Ternary Complex Formation—We next analyzed the regions of ICN needed for assembly of ternary ICN·MAML1·CSL complexes on DNA using an electrophoretic mobility shift assay (Fig. 5*A*). In these experiments, we prepared full-length CSL (by IVT) and a purified recombinant human MAML1 polypeptide (Leu-13—His-74, MAML1-(13–74); Figs. 1*A* and 3*A*) that constitutes the minimal MAML1 region necessary to form complexes with ICN and CSL (18). RAMANK and RAM form stable complexes with CSL on DNA (Fig. 5*A*, *lanes 3* and *5*), but ANK does not (*lane 7*). Although the MAML1-(13–74) does not form stable complexes with CSL and DNA alone (Fig. 5*A*, *lane 2*), the addition of the MAML1-(13–74) to the complex of RAMANK with CSL and DNA causes an additional supershift, indicating formation of a ternary complex (Fig. 5*A*, *lane 4*). ANK also reproducibly forms a weak complex with MAML1-(13–74), CSL, and DNA (Fig. 5*A*, *lane 8*), a finding consistent with reports from other groups (17, 20). ANK is required for recruitment of MAML1 to the multiprotein-DNA complex, as a supershift is not observed upon addition of MAML1-(13–74) (Fig. 5*A*, *lane 6*) or GST-MAML1-(13–74) (data not shown) to RAM·CSL·DNA complexes.

Given the requirement of ANK for recruitment of MAML1-(13–74) and the lack of stable interaction between CSL and MAML1-(13–74) without ICN proteins, we tested whether either RAMANK or ANK binds stably to MAML1 (13–74) in the absence of other factors. Purified recombinant MAML1-(13–74) did not bind GST-RAMANK or GST-ANK (Fig. 5*B*), indicating that recruitment of the MAML1 polypeptide into a multiprotein-DNA complex requires the presence of both ICN and CSL.

We then identified the regions of CSL required to confer binding of MAML1-(13–74) in the presence of RAMANK and DNA. Although CSL-(31–435) binds DNA and assembles complexes with RAMANK, it fails to assemble ternary complexes with MAML1-(13–74) (Fig. 5*C*). Further studies showed that sequences both N- and C-terminal to the DNA-binding region of CSL are necessary, because complex formation is partially (CSL-(31–452)) or fully (CSL-(31–435)) abrogated when the N- and C-terminal extensions are deleted (Fig. 5*C*).

DISCUSSION

Understanding how ICN, CSL, and MAML1 associate with one another and with DNA is central to clarifying how Notch signaling activates transcription. The primary goal of these studies was therefore to define the regions of ICN, CSL, and MAML1 required for formation of complexes with one another and with DNA. This work also has implications for understanding the pathophysiology associated with expression of consti-

tutively active forms of ICN in tumor cells, which harbor complexes of ICN with endogenous CSL and MAML1 in the nucleus.

A Model for the ICN Transcriptional Activation Complex—The combination of biochemistry and bioinformatics used in these studies supports a new model for the architecture of the ICN-dependent transcriptional activation complex (Fig. 6). The primary features of the model are that: (i) the CSL transcription factor is structurally related to proteins in the Rel homology family of transcription factors; and (ii) ICN and CSL combine to create the MAML-1 recruitment site.

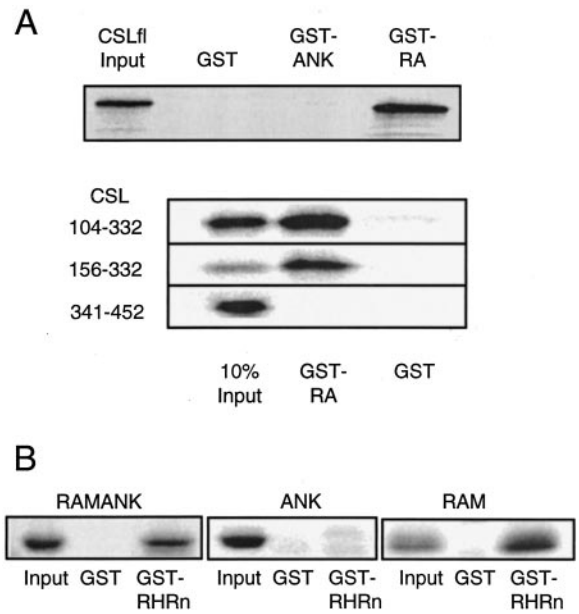

FIG. 4. **Requirements for formation of CSL·ICN complexes.** *A*, glutathione beads with bound GST-RAMANK, GST-ANK, or GST were incubated with various forms of ^{35}S-labeled CSL prepared by IVT. Bound CSL was then detected by autoradiography after SDS-PAGE. The faint bands of lower molecular weight in the *upper panel* likely represent truncated translation products from the IVT reaction. *B*, glutathione beads with bound GST-CSL-(156–332) or GST alone were incubated with RAMANK, ANK, or RAM as indicated. Bound ICN proteins were released by heating in SDS buffer, resolved by SDS-PAGE, and then detected by staining with Coomassie Blue. Lanes marked *Input* contain 20% of the total ICN protein used for each pull-down.

The prediction that CSL proteins are structurally related to the Rel homology family of transcription factors emerged from analysis of the CSL and Su(H) amino acid sequences using the 3D-PSSM structure-prediction server. The Rel homology family of DNA-binding proteins encompasses a wide range of transcription factors including the Rel/NF-κB proteins, NFAT proteins, and osmotic stress-response proteins such as TonEBP (27). Rel homology regions in these transcription factors typically span roughly 250–275 amino acids and consist of two immunoglobulin-like domains (RHR-n and RHR-c) connected by a short linker sequence. On the basis of the 3D-PSSM prediction, the RHR-n domain of human CSL encompasses residues 159–328, and the RHR-c domain spans residues 341–433.

CSL Structure-Function Relationships Echo the RHR Domain Boundaries Identified by 3D-PSSM—Several experimental findings support the 3D-PSSM prediction that CSL contains both RHR-n and RHR-c structural units. Most importantly, the complete RHR region, as well as the isolated RHR-c domain, contains a predominance of β-stranded secondary structure and little α-helix. Furthermore, our studies confirm that a construct encompassing only the isolated RHR-n domain constitutes the high affinity binding site for RAM. In addition, the C-terminal end of the region of CSL required for high affinity binding of cognate DNA coincides almost precisely with the C-terminal end of the predicted RHR-c domain.

Our mixing studies using different-sized CSL variants are consistent with previous conclusions that the functional DNA-binding unit of CSL is a monomer (35). Binding of DNA by CSL requires a remarkably extensive region of the protein (residues 31–435), spanning the entire RHR domain and an additional 125 residues toward the N terminus. This requirement that the entire RHR participate in DNA binding is shared by the other Rel homology transcription factors, which exhibit a variety of

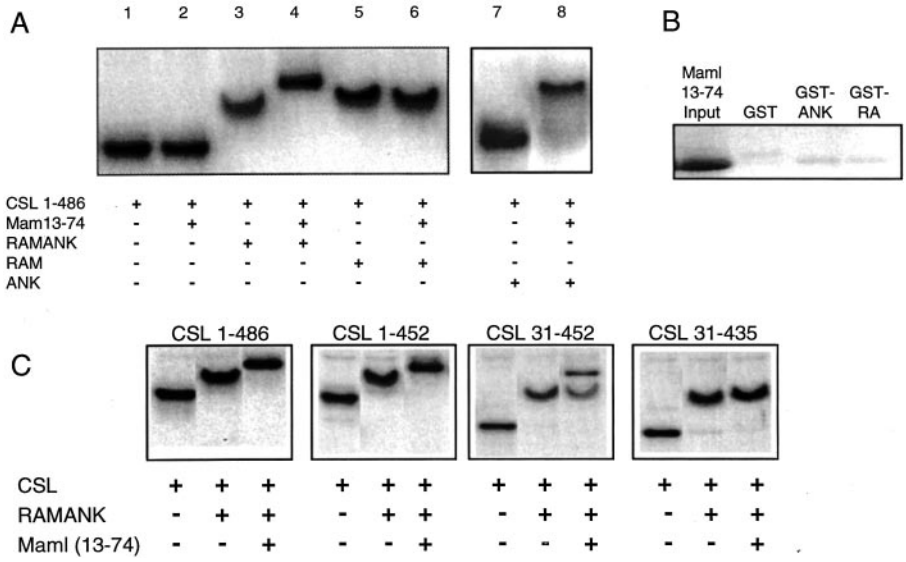

FIG. 5. **Requirements for formation of stable ternary CSL·ICN·MAML1 complexes on DNA.** *A*, EMSA. The indicated combinations of proteins were incubated with a ^{32}P-labeled oligonucleotide with a CSL consensus binding site, resolved by EMSA, and visualized by autoradiography. The concentration of ANK in *lane 8* was 300-fold greater than the concentration of RAMANK used in *lane 3*. *B*, MAML1-(13–74) does not bind to ANK or RAMANK alone. GST, GST-ANK, and GST-RAMANK were each captured on glutathione beads and then incubated with purified recombinant MAML1-(13–74). Proteins bound to the beads were released by heating in SDS buffer, resolved by SDS-PAGE, and then detected by staining with Coomassie Blue. *C*, sequences both N- and C-terminal to the DNA binding region of CSL are necessary for ternary complex formation on DNA. Each panel represents a series of EMSAs performed as described in *A* with the indicated CSL variant and DNA alone (*left*), DNA and RAMANK (*center*), or DNA, RAMANK, and MAML1-(13–74) (*right*).

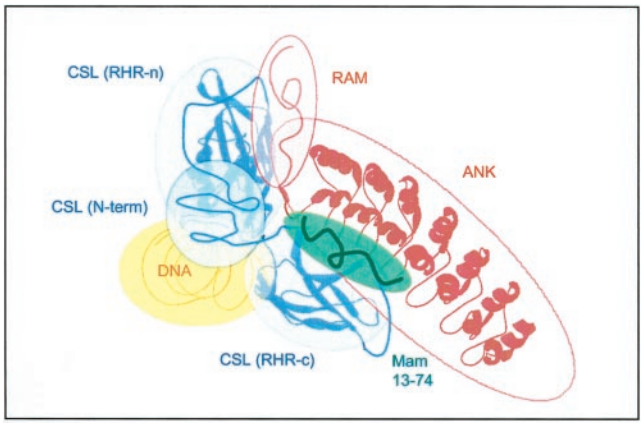

FIG. 6. **Model for the organization of the CSL·ICN·MAML1 ternary complex on DNA.** The drawing illustrates contacts identified experimentally by mapping the interactions of CSL-(1–452) with DNA, RAMANK, and MAML1-(13–74). The structures used to create ribbon diagrams for the RHR domain within CSL and the ankyrin repeat domain within ICN are from the structures of chicken c-*rel* on DNA and human Bcl-3, respectively. The N-terminal region of CSL also required for DNA binding is represented as a *sphere* with an unknown fold. Both RAM and MAML1-(13–74) are illustrated schematically as unstructured polypeptides.

structural arrangements in DNA binding complexes. In high resolution structures of complexes between RHR transcription factors and DNA, both the RHR-n and RHR-c domains participate in the DNA binding interface. Although the Rel/NF-κB (31, 32) and TonEBP (34) proteins bind DNA as dimers, with the primary dimerization interface provided by the RHR-c domain, the RHR domain of NFAT1 binds to DNA as a monomer, in which coooperativity with the Fos/Jun heterodimer enables high affinity binding to cognate DNA (38). By making additional contacts with the DNA, the extra N-terminal region of CSL may effectively obviate the need for dimerization (*e.g.* Rel/NF-κB and TonEBP) or for additional factors (*e.g.* cooperativity between NFAT1 and Fos/Jun) to achieve high affinity DNA binding. By binding DNA as a monomer, CSL most closely resembles the NFAT proteins, which also bind to DNA as monomers, cooperating with the Fos/Jun heterodimer to form high affinity protein-DNA complexes.

The RHR-n Domain of CSL Binds Directly to RAMANK—After ICN is translocated into the nucleus, binding of ICN to CSL·DNA complexes is likely to be the first step required for activation of transcription of Notch-responsive genes. Previous studies have reported the existence of a strong CSL binding site within the RAM region, and a weaker binding site in the ANK domain (39).

We thus produced RAM, ANK, and RAMANK recombinantly for functional studies, first characterizing these polypeptides using biophysical methods, because such analysis of these domains from ICN proteins has previously been limited. Our biophysical studies confirm and extend previous studies of the seven-ankyrin repeat region of *Drosophila* Notch (3, 4). Circular dichroism measurements demonstrate that ANK from human Notch1 exhibits a similar amount of helical content and increased thermal stability when compared with the ankyrin repeat region of *Drosophila* Notch. Biophysical characterization by CD further demonstrates that the RAM region, both as an isolated polypeptide and in the context of the adjacent ANK domain, is a flexible polypeptide segment without significant secondary structure. Finally, both ANK and RAMANK remain monomeric at micromolar concentrations, suggesting that multimerization is unlikely to play a role in physiologic Notch activation.

Our data using purified recombinant RAM, ANK, and RAMANK from ICN in biochemical assays demonstrate direct binding of the RHR-n domain of CSL to RAM and RAMANK (Fig. 4). Analysis of recombinant RAM, RAMANK, and ANK by EMSA also indicates that the part of ICN1 responsible for high affinity binding of CSL lies within the RAM region. Taken together, these findings support the conclusion that contacts between the RAM region of ICN1 and the RHR-n domain of CSL mediate tight binding between ICN1 and CSL, even though RAM lacks secondary structure either as an isolated polypeptide or in the context of RAMANK.

The role of the RAM domain has been elusive, as it is not required for transcriptional activation in reporter assays (22), assembly of ICN variants into complexes with CSL in the presence of MAML1 in cells (17), or tumorigenesis (40). All of these assays relied on enforced high level expression of ICN, conditions under which the lower affinity CSL binding site within ANK may suffice for assembly of transcriptional activation complexes. In support of this possibility, high concentrations of ANK alone supershift CSL·DNA complexes in the presence of MAML-1 (Fig. 5*A*), and prior *in vivo* observations indicate that the RAM region is not absolutely necessary for formation of complexes in transformed cells with increased levels of ICN (17). However, physiologically relevant levels of ICNs produced through ligand-stimulated Notch activation are much lower, typically being below the limits of immunologic or biochemical detection. At these levels, the high affinity interaction between RAM and RHR-n may take on much greater importance.

The complex between RAMANK and CSL exhibits both similarities and differences when compared with the complex between IκB and NF-κB, the only example of a complex between ANK and RHR proteins for which high resolution structural information is available (41, 42). In the NF-κB·IκB complex, the interface between the IκB ankyrin repeats and NF-κB encompasses both RHR-c domains of the NF-κB heterodimer. Because CSL binds to DNA as a monomer, it does not present as extensive an RHR-c interface for binding to the ANK domain of ICN1. Contacts between the ANK domain of ICN and the RHR region of CSL thus appear to confer substantially weaker binding affinity (perhaps because only half of the interface seen in the IκB·NF-κB complex would be present between ANK and RHR-c of CSL), and ICN uses the strong binding affinity of RAM for the RHR-n domain as an anchor to stabilize complexes of ICN with CSL.

Recruitment of MAML1 Requires Cooperation between ANK and Regions Near the N- and C-Termini of CSL—Strong evidence now supports the conclusion that association of ICN1 with the CSL·DNA complex facilitates further recruitment of the protein MAML1, which acts as a transcriptional coactivator. In a modified yeast two-hybrid screen, MAML1 interacts with CSL only in the presence of co-expressed ICN (19). In addition, MAML1 co-purifies with CSL from tumor cells only when ICN is also present (17). Our EMSA results show directly that the RAMANK domain of ICN1 recruits MAML1-(13–74) to the CSL·DNA·ICN complex and that high concentrations of the ANK domain alone suffice for this activity. Remarkably, ANK does not appear to function as a simple adaptor that binds independently to MAML1-(13–74) with high affinity, because we could not detect direct interaction between MAML1 and either ANK or RAMANK using highly purified components. Given that MAML1 does not interact strongly with CSL unless ICN is present (a finding confirmed in our EMSA results), this observation suggests that both ANK and CSL cooperate to provide a high affinity binding site for MAML1.

Further support for a model in which CSL, in addition to ICN, is intimately involved in recruiting MAML1 to complexes on DNA comes from our demonstration by EMSA that CSL

proteins truncated at either the N or C terminus fail to recruit MAML1-(13–74) even though they form stable complexes with RAMANK and DNA. The ability of various CSLs to recruit MAML1-(13–74) also correlated tightly with their capacity to form complexes in cells with full length proteins (data not shown). Reasonable predictions for how CSL and RAMANK act cooperatively to bind MAML1-(13–74) are that: (i) the N- and C-terminal regions of CSL are induced to contact MAML1 (13–74) directly only in the context of ANK, or (ii) both RAMANK and the CSL termini partially contribute to MAML1-(13–74) binding. A third possibility, that the presence of the CSL N- and C-termini is required to induce a structural rearrangement of ANK enabling high affinity association between ANK and MAML1-(13–74), is less likely because the ANK domain is well structured as an isolated domain in solution. Clearly, a high resolution structure of the complex will clarify crucial details about the arrangement of the protein components in the transcriptional activation complex.

Because both CSL and RAMANK combine to recruit MAML1, an intrinsic control mechanism exists to prevent premature activation of Notch target genes and to ensure that off-pathway genes are unaffected when Notch becomes activated. First, because MAML1 cannot bind to CSL sites in the absence of activated Notch, target genes downstream of CSL sites continue to be repressed by co-repressors bound to CSL. Moreover, because ICN1 and MAML1 do not form a stable complex without CSL, cross-reactivity of ICN with DNA-binding proteins other than CSL will not inadvertently cause expression of "off-target" gene products. Finally, recent evidence suggests that MAML1 may not only serve as an effector for transcription of Notch target genes but may also simultaneously commit the active complex for subsequent proteasomal degradation (20). By coordinating the timing of Notch activation with respect to its subsequent destruction, MAML1 would thus also enable a mechanism for tight control of the duration and timing of events driven by activated Notch.

Acknowledgment—We thank Vytas Patriub for excellent technical support.

REFERENCES

1. Artavanis-Tsakonas, S., Rand, M. D., and Lake, R. J. (1999) *Science* **284,** 770–776
2. Nam, Y., Aster, J. C., and Blacklow, S. C. (2002) *Curr. Opin. Chem. Biol.* **6,** 501–509
3. Zweifel, M. E., and Barrick, D. (2001) *Biochemistry* **40,** 14344–14356
4. Zweifel, M. E., and Barrick, D. (2001) *Biochemistry* **40,** 14357–14367
5. Fortini, M. E., and Artavanis-Tsakonas, S. (1994) *Cell* **79,** 273–282
6. Jarriault, S., Brou, C., Logeat, F., Schroeter, E. H., Kopan, R., and Israel, A. (1995) *Nature* **377,** 355–358
7. Christensen, S., Kodoyianni, V., Bosenberg, M., Friedman, L., and Kimble, J. (1996) *Development* **122,** 1373–1383
8. Hsieh, J. J., Zhou, S., Chen, L., Young, D. B., and Hayward, S. D. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 23–28
9. Kao, H. Y., Ordentlich, P., Koyano-Nakagawa, N., Tang, Z., Downes, M., Kintner, C. R., Evans, R. M., and Kadesch, T. (1998) *Genes Dev.* **12,** 2269–2277
10. Taniguchi, Y., Furukawa, T., Tun, T., Han, H., and Honjo, T. (1998) *Mol. Cell. Biol.* **18,** 644–654
11. Petcherski, A. G., and Kimble, J. (2000) *Nature* **405,** 364–368
12. Wu, L., Aster, J. C., Blacklow, S. C., Lake, R., Artavanis-Tsakonas, S., and Griffin, J. D. (2000) *Nat. Genet.* **26,** 484–489
13. Schweisguth, F., and Posakony, J. W. (1992) *Cell* **69,** 1199–1212
14. Yedvobnick, B., Smoller, D., Young, P., and Mills, D. (1988) *Genetics* **118,** 483–497
15. Artavanis-Tsakonas, S., Muskavitch, M. A., and Yedvobnick, B. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 1977–1981
16. Aster, J., Pear, W., Hasserjian, R., Erba, H., Davi, F., Luo, B., Scott, M., Baltimore, D., and Sklar, J. (1994) *Cold Spring Harbor Symp. Quant. Biol.* **59,** 125–136
17. Jeffries, S., Robbins, D. J., and Capobianco, A. J. (2002) *Mol. Cell. Biol.* **22,** 3927–3941
18. Weng, A. P., Nam, Y., Wolfe, M. S., Pear, W. S., Griffin, J. D., Blacklow, S. C., and Aster, J. C. (2003) *Mol. Cell. Biol.* **23,** 655–664
19. Petcherski, A. G., and Kimble, J. (2000) *Curr. Biol.* **10,** R471–R473
20. Fryer, C. J., Lamar, E., Turbachova, I., Kintner, C., and Jones, K. A. (2002) *Genes Dev.* **16,** 1397–1411
21. Wallberg, A. E., Pedersen, K., Lendahl, U., and Roeder, R. G. (2002) *Mol. Cell. Biol.* **22,** 7812–7819
22. Aster, J. C., Robertson, E. S., Hasserjian, R. P., Turner, J. R., Kieff, E., and Sklar, J. (1997) *J. Biol. Chem.* **272,** 11336–11343
23. Rost, B., and Sander, C. (1993) *J. Mol. Biol.* **232,** 584–599
24. Rost, B., and Sander, C. (1994) *Proteins* **19,** 55–72
25. Fischer, D., Barret, C., Bryson, K., Elofsson, A., Godzik, A., Jones, D., Karplus, K. J., Kelley, L. A., MacCallum, R. M., Pawowski, K., Rost, B., Rychlewski, L., and Sternberg, M. (1999) *Proteins* **3,** 209–217
26. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2000) *J. Mol. Biol.* **299,** 499–520
27. Edelhoch, H. (1967) *Biochemistry* **6,** 1948–1954
28. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, J. C., eds) pp. 90–125, Royal Society of Chemistry, Cambridge, England, UK
29. Tun, T., Hamaguchi, Y., Matsunami, N., Furukawa, T., Honjo, T., and Kawaichi, M. (1994) *Nucleic Acids Res.* **22,** 965–971
30. Ghosh, S., May, M. J., and Kopp, E. B. (1998) *Annu. Rev. Immunol.* **16,** 225–260
31. Muller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L., and Harrison, S. C. (1995) *Nature* **373,** 311–317
32. Ghosh, G., van Duyne, G., Ghosh, S., and Sigler, P. B. (1995) *Nature* **373,** 303–310
33. Minoguchi, S., Taniguchi, Y., Kato, H., Okazaki, T., Strobl, L. J., Zimber-Strobl, U., Bornkamm, G. W., and Honjo, T. (1997) *Mol. Cell. Biol.* **17,** 2679–2687
34. Stroud, J. C., Lopez-Rodriguez, C., Rao, A., and Chen, L. (2002) *Nat. Struct. Biol.* **9,** 90–94
35. Chung, C. N., Hamaguchi, Y., Honjo, T., and Kawaichi, M. (1994) *Nucleic Acids Res.* **22,** 2938–2944
36. Michel, F., Soler-Lopez, M., Petosa, C., Cramer, P., Siebenlist, U., and Muller, C. W. (2001) *EMBO J.* **20,** 6180–6190
37. Hsieh, J. J., Henkel, T., Salmon, P., Robey, E., Peterson, M. G., and Hayward, S. D. (1996) *Mol. Cell. Biol.* **16,** 952–959
38. Chen, L., Glover, J. N., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) *Nature* **392,** 42–48
39. Kato, H., Taniguchi, Y., Kurooka, H., Minoguchi, S., Sakai, T., Nomura-Okazaki, S., Tamura, K., and Honjo, T. (1997) *Development* **124,** 4133–4141
40. Aster, J. C., Xu, L., Karnell, F. G., Patriub, V., Pui, J. C., and Pear, W. S. (2000) *Mol. Cell. Biol.* **20,** 7505–7515
41. Huxford, T., Huang, D. B., Malek, S., and Ghosh, G. (1998) *Cell* **95,** 759–770
42. Chen, F. E., Huang, D. B., Chen, Y. Q., and Ghosh, G. (1998) *Nature* **391,** 410–413
43. Sreerama, N., and Woody, R. W. (2000) *Anal. Biochem.* **287,** 252–260