Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor

Kathryn P. Sherry, Rahul K. Das, Rohit V. Pappu, and Doug Barrick

Intrinsically disordered regions (IDRs) play important roles in proteins that regulate gene expression. A prominent example is the intracellular domain of the Notch receptor (NICD), which regulates the transcription of Notch-responsive genes. The NICD sequence includes an intrinsically disordered RAM region and a conserved ankyrin (ANK) domain. The 111-residue RAM region mediates bivalent interactions of NICD with the transcription factor CSL. Although the sequence of RAM is poorly conserved, the linear patterning of oppositely charged residues shows minimal variation. The conformational properties of polyampholytic IDRs are governed as much by linear charge patterning as by overall charge content. Here, we used sequence design to assess how changing the charge patterning within RAM affects its conformational properties, the affinity of NICD to CSL, and Notch transcriptional activity. Increased segregation of oppositely charged residues leads to linear decreases in the global dimensions of RAM and decreases the affinity of a construct including a C-terminal ANK domain (RAMANK) for CSL. Increasing charge segregation from WT RAM sharply decreases transcriptional activation for all permutants. Activation also decreases for some, but not all, permutants with low charge segregation, although there is considerable variation. Our results suggest that the RAM linker is more than a passive tether, contributing local and/or long-range sequence features that modulate interactions within NICD and with downstream components of the Notch pathway. We propose that sequence features within IDRs have evolved to ensure an optimal balance of sequence-encoded conformational properties, interaction strengths, and cellular activities.

Notch signaling | intrinsically disordered proteins | sequence design | transcriptional activation | ankyrin repeats

Intrinsically disordered proteins (IDPs) are abundant in eukaryotic proteomes and they participate in a variety of cellular processes such as cell signaling and cell-cycle regulation (1, 2). As autonomous units, IDPs are disordered under physiological conditions. Archetypal IDPs are deficient in hydrophobic residues and are enriched in either polar or charged residues (3). Conformational heterogeneity is a defining characteristic of IDPs (4), and recent studies have shown that IDPs can be partitioned into distinct conformational classes based on sequence features that determine their overall sizes, shapes, secondary structure biases, and amplitudes of fluctuations (5). The heuristics underlying sequence-ensemble-function relationships for IDPs are akin to sequence-structure-function rules that have been uncovered for proteins that spontaneously fold as autonomous units (1, 6).

The global sizes, shapes, and amplitudes of conformational fluctuations of IDPs are controlled by the fraction of charged residues (FCR) and proline contents within IDP sequences (5, 7–9). Charged residues modulate chain compaction through favorable solvation and a balance of intrachain electrostatic attractions and repulsions (7, 8, 10). To first order, the global dimensions of IDPs, often quantified in terms of ensemble averaged radii of gyration (Rg) and the closely related hydrodynamic radii (Rh), are governed by net charge per residue (7, 8, 10) (NCPR, defined as f− + f+, where f− and f+ are the fractions of positively charged and negatively charged residues, respectively). IDPs with NCPR > 0.25 adopt expanded coil-like conformations in which Rs scales as N1/3 (10).

In many IDP sequences (including the RAM region of the Notch receptor) the FCR is greater than 0.3 (11), and the numbers of positively and negatively charged residues are nearly equivalent (low NCPR) (5). For these polyampholytes, global size, shape, and the amplitudes of conformational fluctuations depend on both the composition of oppositely charged residues and on the degree of linear mixing versus segregation of oppositely charged residues (12). Charge patterning can be quantified using an empirical parameter designated as kappa (κ) (12). In IDP sequences with κ values close to zero, oppositely charged residues are well-mixed and allow for chain expansion through favorable solvation and a balance of repulsive and attractive electrostatic interactions. In IDP sequences with κ values close to one, oppositely charged residues are segregated, which drives chain

Significance

Charge patterning is a key feature of intrinsically disordered protein regions. Here we test whether charge patterning is important for biochemical and biological function, using the “RAM” disordered region of the Notch receptor. The Notch signaling pathway is important in stem-cell biology and cancer. Using computer design, we built 13 charge permutants that span a broad range of charge segregation. These permutants have profound effects on conformational properties, binding affinity to the downstream transcription factor, CSL, and potency in transcriptional activation. WT Notch has the optimal segregation value for activation, whereas higher levels of segregation disrupt binding and activation. Our study paves the way for control of biological function through redesign of charge patterning.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. Published under the PNAS license.

Present address: Ra Pharmaceuticals, Cambridge, MA 02140.

Present address: GNS Healthcare, Cambridge, MA 02139.

To whom correspondence may be addressed. Email: pappu@wustl.edu or barrick@jhu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706083114/-/DCSupplemental.
compaction due to attractive long-range interactions between oppositely charged sequence blocks. Whereas the value of $\kappa$ for an IDP sequence predicts whether the patterning of charge interactions will enhance compaction or expansion, FCR determines the magnitude of these effects (5).

In this study we use an intrinsically disordered region (IDR), the RAM region, from the Notch receptor to investigate sequence-ensemble-function relationships that are governed by polyampholytic IDRs (13). The Notch pathway is a transmembrane cell-to-cell signaling pathway that regulates cellular differentiation and stem cell fate determination in metazoans (14–16). The Notch Intracellular Domain (NICD) includes the RAM region and an adjacent ankyrin (ANK) repeat domain (Fig. L4). During Notch activation, ternary complexes containing NICD, the transcription factor CSL (CBF-1/RBP-J, Su(H), Lag-1), and the coactivator Mastermind (MAML) assemble at the promoter regions of Notch target genes to initiate transcription and determine cell fate (17) (Fig. L4). Both the ANK domain and the N terminus of the RAM region of NICD bind to CSL at spatially distinct sites, forming a bivalent complex connected by the 111-residue RAM linker (18). This disordered linker is thought to affect transcriptional activation by enhancing binding-site occupancy on CSL through an effective concentration mechanism (13, 19). In this model, binding of the N-terminal RAM binding motif to CSL leads to an increase in the local/effective concentration of the ANK domain around its binding site. Therefore, quantities such as the mean $R_g$ or closely related quantities such as the mean end-to-end distance of the RAM region are likely to determine the concentration of the ANK domain around its binding site on CSL. It has also been shown that in addition two regions of local structural bias within the linker have sequence-specific effects on transcriptional activation (13). Along with intrachain electrostatic interactions, these nonrandom elements may influence RAM dimensions and affect transcriptional activation by (i) modulating the effective concentration of ANK near its binding site on CSL, (ii) modulating the affinity of NICD to CSL, or (iii) influencing the recruitment of MAML and occupancy at the transcriptional site through linkage effects.

Most of the attention to date has focused on assessments of the effective concentration model for describing how the RAM linker region influences the activation of Notch genes (13, 19, 20). However, the sequence of the RAM region suggests that it is more than just a passive worm-like chain (13). Over a third of the residues within the RAM region are charged (FCR = 0.35). This region is polyampholytic (NCPR = −0.06), with reasonable segregation opposite charges ($\kappa$ = 0.3173). Previous hydrodynamic measurements showed that the RAM region is more compact than expected for a generic random coil of its size. Moreover, the hydrodynamic size of RAM increases as a function of increased salt concentration (13). This behavior is expected for a polyampholyte with segregated blocks of charge, where favorable intrachain electrostatic attractions are screened at higher salt concentrations. The global conformational preferences and sequence features of the RAM region suggest that the dimensions of RAM and the amplitudes of conformational fluctuations of this IDR may be important for activation of the Notch pathway in ways that extend beyond the effective concentration model.
Here, we use the polyampholytic intrinsically disordered RAM region of the NICD as an archetypal system to understand how designed changes to charge patterning impact chain compaction, binding of NICD to CSL, and transcriptional activation. We generated charge permutant sequences of RAM that span a wide range of $\kappa$ values. Using these permutants we measured the effects of charge patterning on the RAM conformational ensemble, on thermodynamics of binding to CSL, and on transcriptional activation. We find that charge patterning within the linker significantly affects RAM compaction, binding affinity of NICD for CSL, and Notch transcriptional activation. Our observations have broader implications for the evolution of sequence-ensemble-function relationships in IDRs and the design of disordered regions with precise functional consequences. To first order, we find that sequence variants that have dimensions similar to the WT RAM region also give rise to similar transcriptional activities. We also find that NICD sequences with RAM regions that are highly compact bind poorly to CSL and significantly reduce the overall transcriptional activity. However, there are exceptions to these general trends, suggesting specific interactions to CSL that are engendered by differences in charge patterning within the RAM region and contributions of sequence features within the RAM region that affect the recruitment of MAML.

**Results**

**Sequences of RAM Charge Permutants.** Analysis of RAM sequences across orthologs of different Notch genes shows that the disordered RAM region is poorly conserved, especially in comparison with the folded ANK domain (19). Despite poor sequence conservation, the RAM $\kappa$ values for Notch1 orthologs are limited to a narrow range between 0.26 and 0.33 (Fig. 2A). These $\kappa$ values are higher than one would expect from a distribution of random sequences that have the same amino acid composition as the RAM sequence from human Notch1 (11) (Fig. 2 B and C).

To determine if the limited dispersion of $\kappa$ values reflects a functional requirement for achieving an optimal balance of sequence-encoded conformational features, CSL binding, and transcriptional activation, we generated charge permutants of the human Notch1 RAM linker sequence (Fig. 1B). These permutants have identical (i.e., WT) sequence compositions. As a result, the permutants have the same NCPR and FCR as WT RAM, but the positions of charged residues are shuffled to increase or decrease charge segregation compared with the WT sequence (13). The WT sequence is fixed at positions 3–29 and 100–132 (residues 1758–1784 and 1856–1888 of human Notch1) in all permutants to avoid disrupting nuclear localization sequences and the interaction between the N terminus of RAM and CSL. Adhering to these design criteria, we created two sets of permutants. In one series, we shuffled the positions of all residues (charged and uncharged) from residue 31 to residue 99. This series of nine permutants spans a wide range of $\kappa$ values. In a second series, only positively and negatively charged residues were shuffled. Although this series of five permutants (designated with asterisks, Fig. 1B) span a narrower range of $\kappa$ values, it tests the effects of charge patterning in an otherwise fixed WT sequence background. The two permutant series combine to give 13 permutants that have very different patterns of oppositely charged residues and span a broad range of $\kappa$ values from 0.1595 to 0.7481, going well above and below the observed ranges of $\kappa$ values for RAM sequences in different Notch genes.

**Charge Patterning Affects RAM Compaction.** We used all-atom simulations and hydrodynamic measurements to characterize the conformational properties of the RAM charge permutants. We performed all-atom thermal replica exchange Monte Carlo simulations using the ABSINTH implicit solvation model and forcefield paradigm (21). This combination of sampling approach and forcefield has yielded verifiably accurate descriptions of conformational ensembles for a range of IDPs (5, 22–27). For all permutants, including the WT sequence, the simulations yield...
coil-like ensembles with heterogeneous degrees of compaction along the sequence and large fluctuations in sizes and shapes. The simulation temperature determines the quality of conformational sampling. Since the overall coil-like nature of ensembles does not change significantly with temperature, we used a recently published analysis method (27) to identify the lowest simulation temperature (338 K) that yielded robust, statistically converged, and reproducible conformational distributions for all permutants and selected these ensembles for structural analysis.

The calculated ensemble-averaged \( R_g \) values for the permutants decrease roughly linearly from \( \sim 34 \, \text{Å} \) to \( 24 \, \text{Å} \) as \( \kappa \) increases (Fig. 3A and Table S1). The permutants are more compact than expected for an equivalent self-avoiding random coil (the same chain-length and sequence of composition as RAM, \( R_g \sim 48 \, \text{Å} \)) by a factor of 1.5 or more and are more expanded than expected for an equivalent compact globule (\( R_g \sim 15 \, \text{Å} \)) by a factor of 1.7 or more. The \( R_g \) values of the RAM permutants bracket the value expected for a Flory random coil (\( \sim 29 \, \text{Å} \)). In this limiting model, the effects of intrachain interactions (repulsions or attractions) are counterbalanced by chain–solvent interactions.

To test the predicted increase in compaction (decrease in \( R_g \)) with increasing \( \kappa \) value predicted by simulation, we measured sedimentation coefficients for a subset of the RAM charge permutants using analytical ultracentrifugation (AUC; Fig. 3B and Table S2). Sedimentation coefficients are influenced by the molecular weight and frictional properties and can be converted to \( R_g \) by combining Stokes’s Law with the Svedberg equation (28). The \( R_H \) values determined by AUC confirm the prediction that chain dimensions decrease as \( \kappa \) increases (Fig. 3B). As with \( R_g \) values from simulations, we see an overall decrease in \( R_H \) with increasing \( \kappa \) values for the subset of permutants we could measure (high-\( \kappa \) RAM permutants failed to express in *Escherichia coli*). The ratio of calculated \( R_g \) to measured \( R_H \) values for this subset of permutants confirms that the RAM permutants have average conformations consistent with the Flory random coil (Fig. 3C). Modest deviations above or below the Flory value are likely the result of stronger sequence-encoded intrachain repulsions or attractions. The relationship between RAM charge patterning and compaction is consistent with previous work showing that RAM compaction is partially screened at high concentrations of monovalent salt (13). Values of \( R_H \) for RAM have a linear dependence on salt concentration and an approximately linear dependence on \( \kappa \).

To obtain a more detailed picture of how charge patterning influences chain conformations within the RAM ensemble we calculated asphericities from simulations of each permutant. The ensemble-averaged asphericity quantifies the degree to which the envelope of conformations can be modeled as a perfect sphere. Polymeric globules have asphericity values between 0.1 and 0.2. For WT RAM and sequence permutants we calculate asphericity values in the range of 0.45 and 0.65, which is consistent with overall prolate ellipsoidal shapes. Although asphericity decreases with increasing \( \kappa \) values (Fig. 3D), the correlation between asphericity and \( \kappa \) is not as strong as that between \( R_g \) and \( \kappa \) (Fig. 3A).

To further explore the relationship between asphericity and chain compaction we generated 2D \( R_g \)-asphericity histograms from our simulated conformational ensembles. These histograms quantify the degree of conformational heterogeneity and reveal correlated changes to sizes and shapes within the ensemble. This is shown for the WT sequence in Fig. 3E. For the RAM permutant ensembles we calculated the distance \( \omega \) between the distributions for permutants and that of the WT sequence. The
quantity \( 1 - \omega \) is defined as the overlap fraction and is plotted in Fig. 3F. The overlap in global conformational properties is high for permutants with near WT \( \kappa \) values (0.2 < \( \kappa < 0.4 \)). The increased \( R_g \) and asphericity values for sequences with \( \kappa \) below 0.2 as well as decreased \( R_g \) and asphericity values for sequences with \( \kappa \) above 0.4 lead to decreased overlaps between the conformational distributions of permutant and that of the WT RAM. Overall, the simulations suggest that increasing \( \kappa \) results in a linear decrease in dimensions and nonlinear variations in the overlap of 2D size and shape distributions with respect to the WT sequence.

**Charge Patterning Affects RAM:CSL Interactions.** As \( \kappa \) increases, the sequence variants of RAM have higher densities of intrachain contacts (Fig. S2). This is a direct consequence of the linear segregation of oppositely charged residues. To determine if changes to RAM conformations that result from changes to charge patterning have an impact on the binding of RAM to CSL, we measured the binding affinities of RAM charge permutants for human CSL using isothermal titration calorimetry (ITC). For these experiments, the charge permutants include C-terminal ANK domains, which facilitated expression and purification of all charge permutants from *E. coli*. These constructs are designated as permutants of RAMANK. We also measured binding affinities for several of the RAM charge permutants without C-terminal ANK domains (Fig. S3 and Table S3).

The WT RAM and RAMANK proteins bind to CSL with affinities of 4.7 \( \times \) 10\(^{-8} \) M\(^{-1}\) and 1.1 \( \times \) 10\(^{-8} \) M\(^{-1}\), respectively. The twofold increase in affinity for RAMANK likely reflects a small enhancement of binding from weak association of the ANK domain with CSL (20, 29, 30). Likewise, there is a modest but significant difference (1.8-fold, Table S3) in the association constant (\( K_a \)) for binding to CSL between WT RAM and a peptide spanning the RAM N-terminal CSL binding site (residues 3–29 in the numbering scheme in Fig. 1). Although the corresponding differences in the free energy of binding (\( \Delta G^\circ \)) are less than 1 kcal/mol, there is a systematic increase in affinity due to the addition of the linker and the ANK domain compared with the RAM peptide.

In contrast to the modest differences in free energies of the three WT constructs, there is a large difference in the enthalpy of binding of RAMANK to CSL (−34 kcal/mol; Table S3) compared with RAM and the RAM peptide (−19.0 and −19.3 kcal/mol). This ~15 kcal/mol decrease in binding enthalpy may derive from favorable specific interactions between the ANK domain and its binding site on CSL. The similarity in binding enthalpies of the RAM peptide and larger RAM construct suggests that the favorable enthalpy term for RAMANK binding does not result from new contacts between the RAM linker and CSL, unless those contacts are promoted by covalent connection with the ANK domain. Note that because binding free energies of these three WT constructs are similar the favorable enthalpy term associated with RAMANK binding is almost completely offset by unfavorable entropy of binding. This large entropic penalty is not likely to be the sole result of translational and solvent entropy associated with ANK–CSL binding; rather, it is likely to result from a decrease in conformational heterogeneity of the RAM linker as a result of bivalent association.

By comparing association constants of RAMANK permutants to \( \kappa \) and ensemble-averaged \( R_g \) values we find that binding of RAM to CSL becomes weaker as charge segregation and compaction increase (Fig. 4). Variants P10 through P13 (\( \kappa = 0.5010–0.7481 \)) have the lowest binding affinities for CSL (\( K_a = 2.5–3.1 \times 10^{-7} \) M\(^{-1}\)), the highest \( \kappa \) values, and the most compact conformational ensembles. P1, P2\(^*\), and P4 have the highest affinities for CSL, lowest \( \kappa \) values, and the most expanded conformational ensembles. WT RAMANK has a similarly high binding affinity for CSL; however, it has a more intermediate degree of charge patterning and compaction. The correlation between RAM \( R_g \) and RAMANK:CSL binding affinity within the permutant series is consistent with RAM compaction affecting the accessibility of the RAM or ANK binding sites and causing significant changes in binding site occupancy (discussed below).

As with WT RAMANK binding enthalpy to CSL, the binding enthalpies of RAMANK permutants are consistently lower than those of the corresponding RAM constructs (Fig. 4C). Overall, this decrease is consistent with favorable enthalpy interactions between ANK and CSL. However, there are considerable variations in binding enthalpies within both the RAM permutant series and the larger RAMANK series. Because binding free energies show a smaller variation, this enthalpy variation leads to a large correlated entropy variation among permutants (Fig. 4D), as is typical of macromolecular systems. As it seems unlikely that variations in charge patterning lead to variations in binding enthalpy (and entropy) between ANK and CSL, such variations among permutants may reflect weak, perhaps heterogeneous “fuzzy” (31, 32) ANK-induced interactions between the RAM linker region and CSL.

**Impact of Charge Patterning in the RAM Linker on Notch Transcriptional Activation.** To determine if RAM charge patterning has functional consequences for Notch signaling we substituted the WT RAM region of human Notch1 NICD with the charge permutant sequences and measured the transcription activities of the resulting NICD permutants in HeLa cells (Fig. 5). In general, transcriptional activation shows high sensitivity to charge patterning of the RAM region. This sensitivity is greater than that seen previously for perturbations of residual secondary structure elements within RAM (13).

For permutants with high \( \kappa \) values, transcriptional activation decreases significantly. Notably, transcription activities of P12 (\( \kappa = 0.6022 \)) and P13 (\( \kappa = 0.7481 \)) are only about 10% of WT levels. For permutants with \( \kappa \) values near WT, transcriptional activation is similar to that of WT; however, there is some variation in activity among these permutants. Interestingly, transcription activities of P7\(^*\) (\( \kappa = 0.4065 \)) and P9 (\( \kappa = 0.4450 \)) are 10–20% higher than that of WT. For permutants with low \( \kappa \) values, transcriptional activation is modestly decreased. These decreases range from 88% of WT for P1 (\( \kappa = 0.1595 \)) to 56% of WT for P2\(^*\) (\( \kappa = 0.2084 \)) and P3 and P4 (\( \kappa = 0.2172 \)). Overall, transcriptional activation of the RAM charge permutants is highest when charge patterning is most similar to WT, decreasing sharply for permutants with high \( \kappa \) values. This provides a biophysical rationalization for the limited dispersion of \( \kappa \) values in Notch-1 orthologs.

Charge segregation may affect Notch transcriptional activation either through effects on RAM linker compaction, which would modulate the effective concentration of the ANK domain, through effects on the strength of the RAM:CSL interaction, or through interactions with downstream signaling components such as MAML. It is unlikely that charge segregation significantly affects NICD cellular localization or posttranslational modifications because there are no significant differences between the activation of permutants with WT sequence backgrounds and those with variable background sequences. In addition, there are no correlations between transcriptional activation and the identity of the two known posttranslationally modified sites within the RAM linker (phosphorylation sites S1791 and S1801) (33, 34).

Charge patterning in the RAM linker likely impacts Notch transcriptional activation through a combination of molecular mechanisms. The sharp decrease in transcriptional activation with increasing charge segregation for \( \kappa > 0.4 \) (Fig. 5) is correlated with increasing compaction of RAM (Fig. 3). This may reflect a reduction in RAMANK:CSL interaction by limiting access of ANK (or RAM) to its binding site on CSL. To test the validity of the sequestration hypothesis, Fig. 5 also shows results from a model used to describe the activation data. In this model, we assume that...
RAMANK is in equilibrium between an open conformation that can bind CSL and a closed, binding-incompetent conformation (Materials and Methods). Here, we assume that the free energy difference between the open and closed conformational states depends linearly on $\kappa$. This model accounts for the sharp transition between transcriptional activation and inactivation as $\kappa$ increases, although it does not explain all of the data (Fig. 5). Specifically, there are outliers at low $\kappa$ values (P2*, P3, and P4) that reproduce only 50–60% of the activity of RAMANK with a WT RAM region.

To the extent that the simple sequestration model fits the transcriptional activation profile it implies that compaction limits access of the RAM region and/or the ANK domain to sites on CSL. This may result either from strong compaction of the RAM linker, which lowers effective concentration [consistent with the sharp decrease in activation as a result of linker truncation (13)], or it may be result of direct steric occlusion of the RAM binding site. Inasmuch as the RAMANK:CSL interaction is dominated by the RAM binding sequence engaging with CSL, the effective concentration model does not account for the decreased affinity of the high-permutant RAMANK constructs, nor does occlusion of the ANK domain by the RAM linker. In contrast, direct occlusion of the RAM binding is consistent with the decrease in affinity and activation.

The deviations from the fit in Fig. 5 suggests that there is more to the effect of the RAM region than a compaction/expansion induced modulation of the effective concentration of the ANK domain around the CSL binding sites. There are likely to be at least two other effects in play: (i) the influence of auxiliary contacts between the RAM region and CSL and (ii) the impact of the RAM region on recruiting MAML and hence influencing occupancy of the ternary complex at the transcriptional site.

**Discussion**

We have studied the impact of charge patterning within the RAM region on global dimensions of RAM, the binding of RAMANK to CSL, and the activation of Notch transcription. We find that the RAM region does more than just generate bivalency or modulate the effective concentration of ANK. Our findings suggest that multiple sequence-encoded features of the intrinsically disordered RAM region influence the transcriptional activity of NICD. The dimensions of the RAM region influence the effective concentration of ANK (Fig. 6A–C). The RAM distal binding motif determines the specificity of RAM binding to CSL. Changes to the sequence of the RAM linker clearly have significant impact on the binding profile of RAMANK to CSL (Fig. 4D). Moreover, the inability to explain all of the transcriptional activity data using global dimensions and the entropy–enthalpy profiles suggests that the RAM region also influences either direct interactions with or indirect recruitment of MAML (Fig. 6D and E). Therefore, the sequence-specific features of the RAM region have to be viewed holistically to the transcriptional activity that is controlled by NICD.

Previous studies showed that changes to the RAM linker length modulate Notch transcriptional activation with a profile consistent with a RAM-mediated concentration enhancement of RAMANK binding site sequence for CSL. RAMANK:CSL binding affinities decrease as RAM charge patterning and compaction increases. (C) Binding enthalpies for WT and RAM (upper) and RAMANK constructs (lower) and for permutant RAM and RAMANK constructs. The binding enthalpy of the 27-residue is indicated with a dashed vertical line. WT RAM and RAMANK $\Delta H^0$ values are connected with a solid diagonal line; permutants 3, 5*, 7*, 8*, and 9 are connected with dashed diagonal lines. (D) Enthalpy–entropy correlation for RAMANK binding to CSL. Solid line is the best fit to the data. In C and D permutant colors are as in Fig. 3A.
permutants, which is captured by a simple two-state model (Eqs. 1–3). Modest variations in transcriptional activity are observed for low-κ mutants. Transcription activities are reported relative to WT NICD (flat dashed line) and are the mean of at least three experiments performed in quadruplicate. Error bars are SEs on the mean.

Fig. 5. Transcriptional activities of RAM charge permutant sequences within NICD. WT levels of activation are observed for charge permutants with κ values similar to WT RAM. Significant decreases in activation are observed for high-κ permutants, which is captured by a simple two-state κ-dependent equilibrium between binding-competent and incompetent forms of RAMANK (Eqs. 1–3). Modest variations in transcriptional activity are observed for low-κ permutants. Transcription activities are reported relative to WT NICD (flat dashed line) and are the mean of at least three experiments performed in quadruplicate. Error bars are SEs on the mean.

ANK (13, 19, 20). If highly compact RAM linkers are unable to position the RAM and ANK domains near their binding sites, then distal communication between the RAM peptide and the ANK domain will be reduced (Fig. 6C), and this should in turn diminish transcriptional activation. In addition, direct interactions between permutated RAM-linker and the RAM binding or ANK domains (Fig. 6E) may inhibit binding to CSL binding and transcriptional activation. Our results for P10, P12, and P13 are consistent with these inhibitory mechanisms. Compaction of the RAM region compromises binding of RAMANK to CSL and this weakens transcriptional activation.

For RAM permutants with κ < 0.3, the variable relationship between charge patterning and transcriptional activation suggests a high sensitivity to RAM linker sequence changes among expanded conformers. Although these permutants adopt expanded global conformations that are consistent with their low κ values, comparisons of intermediate states show local compaction for these permutants compared with WT (Fig. S2). These permutants bind CSL with a affinity similar to WT RAM; however, sequence-specific changes in local structure and intermolecular electrostatic interactions may perturb assembly of downstream interactions such as ANK and MAML1 binding to CSL, thereby affecting transcription. In addition, transcriptional activation of the low-κ permutants may be more sensitive to disruption of specific sequence motifs within the RAM linker that have been shown to affect activation despite being distal from the RAM and ANK binding sites (13).

Interestingly, the highest levels of transcriptional activation are observed for WT Notch and for charge permutants with κ values similar to WT (0.3 < κ < 0.4). Although permutants in this intermediate range have RAMANK:CSL binding affinities that are slightly lower than WT, their transcriptional activities are the same as (or slightly higher than) WT RAMANK, suggesting that ternary complex formation is most favorable for these permutants compared with most of the lower and the higher κ permutants. Because intrachain and interresidue distances are most similar to the WT sequence for intermediate κ permutants (Fig. S2), ternary complex formation is most likely favored by WT charge patterning that maximizes RAM and ANK binding-site accessibility, effective concentration enhancement, and favorable intermolecular electrostatic interactions through a balance of compaction and sequence-specific effects (Fig. 6). Our results therefore suggest a biophysical basis for the observed bounds on κ values for RAM sequences in various Notch genes (Fig. 2).

In this work we used the empirical parameter κ to quantify the extent of segregation or mixing of oppositely charged residues within RAM variants. Other metrics have been developed that take patterning effects along the linear sequence into account. Sawle and Ghosh (35) introduced a parameter referred to as sequence charge density (SCD), which is anticorrelated with κ. Its efficacy for describing polyampholytes has been demonstrated in work on the dimensions of unfolded states as well as the phase behavior of polyampholytes (36). We calculated the SCD parameter for each of the RAM variants. The correlation coefficient between SCD and κ is r = −0.83, which is in accord with the findings of Sawle and Ghosh (35). Although the analytical approaches to calculating SCD and κ are quite different, both approaches provide similar correlations to binding affinity and RAM dimensions.

The relationships among RAM charge patterning, compaction, NICD bivalency, and Notch transcriptional activation demonstrate that charge patterning determines the global conformational preferences of the RAM linker, which in turn determines its interaction profiles and ability to assemble the ternary complex. Clearly, the disordered region plays multiple roles in influencing Notch signaling. Going forward, it would be essential to uncover the auxiliary interactions that define the apparently “fuzzy” interface between RAMANK and CSL. It is also important to uncover how the RAM region might influence the recruitment of MAML and determine the occupancy of the ternary complex at the transcriptional site. These investigations will require a combination of sequence designs introduced in this work, direct measurement of weak/labile interactions in the RAMANK:CSL interface, and determination of the energetic impact of the RAM region on assembly of the ternary complex.

From a practical standpoint, there is growing interest in being able to control Notch signaling for applications in regenerative medicine (37) and reprogramming cardiomyocytes into conductive cells (38). The controlled differentiation of adult stem cells requires the ability to inhibit as well as activate Notch signaling (37). We have demonstrated that sequence-encoded changes to the charge patterning within RAM help in directly modulating the transcriptional output of Notch genes. We propose that our findings might pave the way for using modern genetic engineering tools to design on/off variants of the Notch receptor that work in tandem with one another as desired. Our findings take on additional significance because many polyampholytic IDRs mediate multivalent protein–protein interactions (39–52). Accordingly, charge patterning is likely to be an important evolutionary determinant of sequence-ensemble-function relationships in several polyampholytic IDRs (5, 27). Our approach paves the way for deploying sequence design combined with biophysical studies to uncover the evolutionary determinants of IDR functions and for modulating the outputs of transcriptional circuits and signaling pathways in synthetic biology applications.

Materials and Methods

Protein Expression and Purification. Human CSL was expressed and purified as described previously (13). WT RAMANK, spanning residues 1844–2126 of human Notch1, was expressed as an MBP fusion protein as described previously (13). WT RAM and RAMANK charge permutants were expressed with an N-terminal His tag followed by a tobacco etch virus (TEV) protease cleavage site with a Met residue located in the P1 position. In addition, WT RAM and RAMANK charge permutants were expressed with C-terminal Leu and Glu residues. RAM constructs were expressed in BL21 (DE3*) E. coli cells grown in LB medium to an optical density of 0.8–1. WT RAM and RAM charge permutant expression was induced overnight by
adding 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and lowering the growth temperature to 20 °C RAMANK charge permutant expression was induced for 4 h by adding 1 mM IPTG and maintaining the growth temperature at 37 °C. Bacteria were collected by centrifugation and were stored at −80 °C.

Bacteria were lysed by high-pressure homogenization after resuspending in 25 mM Tris HCl, pH 8.0, 50 mM NaCl, and 0.5 mM Tris(2-carboxyethyl) phosphine (TCEP). WT RAM and RAM charge permutants were clarified by centrifugation immediately after lysis. Lysate supernatants were treated with DNase I and Benzonase for 1 h at room temperature. RAMANK charge permutant lysates were treated with DNase I and Benzonase (Sigma-Aldrich) for 30 min at room temperature and clarified by centrifugation.

After adding NaCl to a concentration of 500 mM, WT RAM and RAM charge permutants were purified from the treated lysates with a bench-top column with Ni-NTA agarose resin equilibrated in 25 mM Tris HCl, pH 8.0, 500 mM NaCl, and 0.5 mM TCEP. Permutants were eluted with 300 mM imidazole and were subsequently dialyzed overnight with TEV protease into anion-exchange buffer (25 mM sodium phosphate buffer, pH 7.0, 100 mM NaCl, and 0.5 mM TCEP). WT RAM and RAM charge permutants were further purified from the dialysate by anion-exchange chromatography on SP Sepharose resin (GE Healthcare Life Sciences).

RAMANK charge permutant lysis pellets were resuspended in resuspension buffer (6 M urea, 25 mM Tris HCl buffer, pH 8.0, 500 mM NaCl, and 0.5 mM TCEP) and were clarified by centrifugation. RAMANK charge permutants were purified from the supernatant with a bench-top column with Ni-NTA agarose resin equilibrated in resuspension buffer. Permutants were eluted as above, dialyzed extensively into 25 mM sodium phosphate buffer, pH 7.5, 50 mM NaCl, and 2 mM β-mercaptoethanol and treated with TEV protease. RAMANK charge permutants were further purified with an additional nickel column step or with anion-exchange chromatography as described above.

**Sedimentation Velocity AUC.** WT RAM and RAM charge permutants were dialyzed into 25 mM sodium phosphate buffer, pH 7.5, 50 mM NaCl, and 0.1 mM TCEP. Dialyzed RAM was combined with dialysis buffer to 100 μM and diluted serially with equivalent volumes of dialysis buffer. Instrument setup and data analysis were described previously (13).

**ITC.** WT RAM, RAM charge permutants, and CSL were dialyzed in 25 mM sodium phosphate buffer, pH 7.0, 150 mM NaCl, and 0.5 mM TCEP. For RAM peptide titrations, peptide (GenScript) was resuspended in buffer of identical composition and used without dialysis. WT RAMANK and RAMANK charge permutants were dialyzed in 25 mM sodium phosphate buffer, pH 7.5, 50 mM NaCl, and 0.5 mM TCEP. Protein concentrations were measured by UV absorbance after purification of dialyzed protein samples through 0.22-μm PVDF membranes.

ITC thermograms were obtained at 25 °C on a MicroCal VP-ITC. For WT RAM, RAM peptide, and RAM charge permutant titrations, 85–100 μM RAM was titrated into 7–10 mM CSL. WT RAMANK and RAMANK charge permutant titrations, 20–40 μM CSL was titrated into 2–4 μM RAMANK. Data were analyzed with a single-site binding model with accompanying Origin software. Fitted parameters are reported as an average of at least three experiments ± SD.

**Transcription Activity Assays.** The NICD construct spans residues 1758–2555 of human Notch1 in a pcDNA3.1(+) vector. For the charge permutants, residues 1758–1888 were substituted with synthetic gene fragments containing the RAM permutant sequences with an N-terminal methionine. HeLa cells were propagated in Dulbecco’s modified Eagle’s medium with 10% (vol/vol) FBS and 1% (vol/vol) penicillin/streptomycin at 37 °C.

Cells were seeded in 24 well-plates 24 h before transfection at a density of 2 × 10^5 cells per well. Cells were transiently transfected with 225 ng TP1-luc reporter plasmid, 75 ng Renilla transfection control plasmid, and 100 ng NICD-expressing plasmid using Lipofectamine LTX (Life Technologies). An empty pcDNA3.1(+) plasmid was used as a negative control for all experiments. Cells were harvested 44–48 h after transfection by lysing in passive lysis buffer and were assayed using dual-luciferase reagent on a GloMax Multi plate luminometer (Promega). Luciferase activities for each construct were measured in quadruplicate for each transfection and the transfactions were repeated at least three times for each construct.

![Diagram](Fig 6. Proposed mechanisms for the effects of RAM charge patterning and compaction on binary and ternary complex formation. A–C show the impact of charge patterning mediated expansion/compaction on the effective concentration (dashed blue circle) of the ANK domain around its binding site on CSL. (A) Well-mixed sequences lead to expanded RAM linkers, reducing the effective concentration of the ANK domain. (B) Moderate charge segregation within the RAM linker leads to chain compaction and can increase the effective concentration of the ANK domain around CSL. (C) High charge segregation within the RAM linker can lead to overcompaction, decreasing the effective concentration of the ANK domain around CSL. D and E show potential additional interactions involving the RAM linker that can promote or inhibit transcriptional activation. (D) Additional interactions between the RAM linker and CSL (and perhaps MAML) that promote transcriptional activation. These secondary interactions may be disrupted by permutation, which would decrease transcriptional activation. (E) New intramolecular interactions between the RAM linker and the ANK and RAM binding sites, which would disrupt interaction with CSL and decrease transcriptional activation.)
Expression of NICD charge permutants was verified with Western blotting (Fig. S4).

Transcriptional activation data were fit using a model that assumes an equilibrium between a binding-competent (RAM) and a binding-incompetent (RAM) state of NICD, with an equilibrium constant

\[ K_a = \frac{\text{RAM}}{\text{RAM}^*} = \exp\left(\frac{-\Delta G}{RT}\right) = \exp\left(-\frac{(\Delta H + \Delta S)k_B}{RT}\right). \]  

[1]

The parameters \(a\) and \(b\) are the free energy modeled at \(x = 0\), and the sensitivity of the free energy difference between RAM and RAM\(\Delta\) to changes in \(x\). Assuming that only RAM can bind CSL, the fraction of CSL bound with RAM is

\[ f_{\text{bound}} = \frac{K_a}{1 + K_a} \frac{\text{RAM}}{\text{RAM}^*} = \frac{K_a}{1 + K_a} \frac{\text{RAM}}{\text{RAM} + \text{RAM}}. \]  

[2]

Here, [RAM] is the total RAM concentration and \(K_a\) is the binding constant of competent RAM to CSL. The transcriptional activity (\(\alpha\)) should be proportional to the fraction of CSL that is bound by NICD, that is,

\[ \alpha = f_{\text{bound}} + A_{\text{free}} f_{\text{free}}. \]  

[3]

where \(f_{\text{free}}\) is the fraction free (1 – \(f_{\text{bound}}\)) and \(A_{\text{bound}}\) and \(A_{\text{free}}\) are the transcriptional activity on NICD associated with the bound and free forms, respectively. Fitting was carried out with four adjustable parameters (\(a, b, K_a\), and \(A_{\text{free}}\)) using weighted nonlinear least squares in Mathematica. Reciprocal square (SDs (Fig. S5) were used as weights in the fit.

All-Atom Simulations. All simulations and analyses were performed using the CAMPARI molecular modeling software suite (campaari.sourceforge.net). In these simulations we used the ABSINTH implicit solvation model, whereas all polypeptide and ion contributions were modeled in atomic detail. The simulations were carried out using spherical boundary conditions. In each simulation, the system comprised the polypeptide chain with Glu and Asp side chains in their anionic (deprotonated) charge states and Lys and Arg side chains in their cationic (protonated) charge states, neutralizing Na\(^+\) and Cl\(^-\). Ions plus 49 excess ion pairs to mimic 10 mM NaCl enclosed within a spherical droplet of radius 125 Å. This droplet radius was chosen to be larger than the end-to-end distance distributions expected for RAM modeled as an atactic self-avoiding random walk, which is referred to as the excluded volume limit in Fig. S5. These simulations revealed that none of the sequences sampled conformations with end-to-end distances larger than 250 Å, thus guarding against artifacts due to confining effects imposed by small droplet size. The cutoff distances for van der Waals interactions and for electrostatic interactions between neutral groups were set to 10 Å and 14 Å, respectively. No cutoffs were used for electrostatic interactions involving Na\(^+\) and Cl\(^-\) ions and the charge-groups of Asp, Glu, Arg, and Lys side chains.

Parameters for all of the interaction terms in the force field were taken from the OPLS3.2 opsite set. This set incorporates ABSINTH-based parameters for van der Waals interactions, torsional potentials, and reference free energies of solution for solution groups. The partial charges and the neutral group paradigm are adapted from the OPLS-AA/L force field (S3). In the ABSINTH paradigm the effects of solvent-mediated interactions are captured using an implicit representation of the solvent that models the contributions from conformation-dependent dielectric inhomogeneities. The effects of mobile ions are simulated using explicit representations of Na\(^+\) and Cl\(^-\) ions (S4). All other details are identical to the simulations of p27 permutants that was recently published by Das et al. (S2).

Conformational space was sampled using Metropolis Monte Carlo (MMC) simulations. The degrees of freedom for these MMC simulations include the backbone torsion angles \(\varphi, \psi\), and \(\omega\), side-chain torsion angles \(\chi\), and the rigid body coordinates of polypeptides and solution ions. The move sets include translation of ions combined with small- and large-scale conformational changes of the polypeptide degrees of freedom. All simulations are achieved through a combination of local, pivot, and concerted moves and their frequencies are based on the decision tree used in previous work (S2).

ACKNOWLEDGMENTS. We thank current and past members of the D.B. and R.V.P. laboratories, specifically Christine Hatem, Scott Johnson, Alex Holehouse, Tyler Harmon, Kirsten Ruff, and Andreas Vitalis, for helpful discussions and technical assistance and the Johns Hopkins University Center for Molecular Biophysics for providing facilities and resources. This work was supported by Grant GM060001 from the National Institutes of Health (to D.B.) and Grants MCB 1121867 and MCB 1614766 from the National Science Foundation (to R.V.P.).