

# The mechanism and high-free-energy transition state of lac repressor–lac operator interaction

Rituparna Sengupta<sup>1,2</sup>, Michael W. Capp<sup>2</sup>, Irina A. Shkel<sup>2,3</sup> and M. Thomas Record, Jr<sup>1,2,3,\*</sup>

<sup>1</sup>Program in Biophysics, University of Wisconsin—Madison, Madison, WI 53706, USA, <sup>2</sup>Department of Biochemistry, University of Wisconsin—Madison, Madison, WI 53706, USA and <sup>3</sup>Department of Chemistry, University of Wisconsin—Madison, Madison, WI 53706, USA

Received August 04, 2017; Revised September 07, 2017; Editorial Decision September 13, 2017; Accepted September 22, 2017

## ABSTRACT

**Significant, otherwise-unavailable information about mechanisms and transition states (TS) of protein folding and binding is obtained from solute effects on rate constants. Here we characterize TS for lac repressor(*R*)–lac operator(*O*) binding by analyzing effects of *RO*-stabilizing and *RO*-destabilizing solutes on association ( $k_a$ ) and dissociation ( $k_d$ ) rate constants. *RO*-destabilizing solutes (urea, KCl) reduce  $k_a$  comparably (urea) or more than (KCl) they increase  $k_d$ , demonstrating that they destabilize TS relative to reactants and *RO*, and that TS exhibits most of the Coulombic interactions between *R* and *O*. Strikingly, three solutes which stabilize *RO* by favoring burial/dehydration of amide oxygens and anionic phosphate oxygens all reduce  $k_d$  without affecting  $k_a$  significantly. The lack of stabilization of TS by these solutes indicates that *O* phosphates remain hydrated in TS and that TS preferentially buries aromatic carbons and amide nitrogens while leaving amide oxygens exposed. In our proposed mechanism, DNA-binding-domains (DBD) of *R* insert in major grooves of *O* pre-TS, forming most Coulombic interactions of *RO* and burying aromatic carbons. Nucleation of hinge helices creates TS, burying sidechain amide nitrogens. Post-TS, hinge helices assemble and the DBD-hinge helix-*O*-DNA module docks on core repressor, partially dehydrating phosphate oxygens and tightening all interfaces to form *RO*.**

## INTRODUCTION

The lac repressor (*R*)–lac operator (*O*) interaction is a classic example of an on–off switch in prokaryotic gene regulation (1–4). Large conformational changes in both repressor and operator occur in forming the highly stable *RO* complex (5–13). These are reversed by binding of feedback-control inducer ligands, dissociating the *RO* complex to al-

low expression of the lac genes (14–17). For an *O* site embedded in polymeric DNA, the kinetics of *RO* formation at low pM concentrations of *R* and *O* were determined and interpreted in terms of facilitated diffusion of *R* on DNA after non-specific binding (18–22). To obtain information about the series of conformational changes in *RO* formation once *R* has located *O* and about the high-free-energy transition state (TS) in this mechanism, studies with small *O*-containing DNA fragments under conditions where diffusion to *O* is not rate-limiting are needed. In a classic study, Whitson and Matthews (23) determined and compared effects of salt and temperature on rate constants for *RO* formation and dissociation on a 40 bp fragment and on polymeric DNA.

A plausible mechanism of *RO* formation can be proposed on the basis of crystallographic (7,8) and nuclear magnetic resonance (NMR) (9–11) structural information. This process starts with insertion of the flexibly-tethered DNA binding domains (DBD) of *R* into the major grooves of *O*, followed by folding of the flexible tethers (hinge regions) into  $\alpha$ -helices and insertion in the central minor groove, bending *O*. Folding of the tethers allows the DBD-hinge helices-*O* DNA assembly to interact with the core of *R* (24,25). While certainly a logical order of interactions and conformational changes, this structural mechanism has not been tested by kinetic studies. Here we report solute effects on association ( $k_a$ ) and dissociation ( $k_d$ ) rate constants, determined under conditions where diffusion of *R* to the *O* site is not rate-determining. Interpretation of these data, using solute-model compound interaction information ( $\alpha$ -values) and molecular modelling, provides information about interfaces and conformational changes involved in forming the high-free-energy transition state (TS) that determines these rate constants, and thereby tests the structure-based proposal for the mechanism of *RO* formation.

A large reduction in water accessible surface areas (ASA) of *R* and *O* (i.e. a large negative  $\Delta$ ASA) occurs in the *RO* interaction. Large  $\Delta$ ASA also occur in most other self-assembly interactions of proteins, nucleic acids and their complexes, as well as in the steps of operation of these assemblies as molecular machines. The ASA reduction in *RO*

\*To whom correspondence should be addressed. Tel: +1 608 262 5332; Fax: +1 608 262 3453; Email: mtrecord@wisc.edu

formation occurs as **R** and **O** functional groups reduce or eliminate their interactions with water and solutes to form interfaces of the **RO** complex. Solute effects on rate constants for **RO** formation and decomposition provide information about the amount and composition of the  $\Delta\text{ASA}$  of formation of the high-free-energy transition state from reactants, just as effects of temperature on these rate constants provides information about the enthalpic barriers for forming **TS** from reactants.

While structural information about **TS** of enzyme-catalyzed reactions has been obtained using stable **TS** analogs (26,27), no similar structural approach is generally available for non-covalent **TS**. To determine what conformational changes occur and what order they occur in, general probes that are sensitive to ASA changes in mechanistic steps are needed. Because solutes affect biopolymer processes by interacting with the biopolymer surface exposed or buried in the process (28), solutes are a natural choice as both thermodynamic and mechanistic (kinetic) probes of ASA changes.

Guinn *et al.* (29) characterized the **TS** for folding of 13 proteins by analyzing effects of urea, GuHCl and temperature on folding and unfolding rate constants. In protein folding and unfolding to **TS**, GuHCl primarily probes changes in accessible surface area ( $\Delta\text{ASA}$ ) of amide groups, while urea probes both amide and hydrocarbon  $\Delta\text{ASA}$  and activation heat capacities primarily probe hydrocarbon  $\Delta\text{ASA}$ . Amounts of amide and hydrocarbon ASA buried in **U**→**TS** and **TS**→**F** for folding all 13 proteins showed that **TS** is advanced (because a majority (50–90%) of the  $\Delta\text{ASA}$  for **U**→**F** is buried in **U**→**TS**), and that the  $\Delta\text{ASA}$  of **U**→**TS** is preferentially amide. These findings are consistent with a folding mechanism in which all elements of unstable 2° structure form early and intervening backbone regions then organize in **TS** to nucleate coalescence of these 2° structure elements post-**TS**. This research served as proof-of-principle of the use of solutes as probes of changes in surface area in biopolymer processes. Solute and salt effects on the thermodynamics and kinetics of RNA polymerase-promoter DNA open complex formation and stabilization were also used to identify and characterize large conformational changes and new interfaces in the key intermediates and in the DNA opening-closing **TS** in transcription initiation (30–32).

Here, we extend this approach to characterize **TS** and propose a mechanism for the structurally well-characterized **RO** interaction. We measure and interpret effects of four solutes (the **RO** destabilizer urea and **RO**-stabilizers glycerol, proline and glycine betaine (GB)) on the association and dissociation rate constants and equilibrium constant of the **RO** interaction. We also report and interpret effects of the stabilizer tetraethylene glycol (TEG) on the **RO** equilibrium constant, and analyze published effects of KCl concentration (23) on **RO** rate and equilibrium constants. Interactions of these solutes with the different coarse-grained types of unified atoms of proteins have been investigated previously, and strengths of interaction ( $\alpha$ -values) with these unified atoms and with anionic phosphate oxygens of DNA have been determined (33–36). Approximately 80% of the predicted  $\Delta\text{ASA}$  of **RO** dissociation (4600 Å<sup>2</sup> of the total 5800 Å<sup>2</sup>) is from exposure of these types of unified atoms. Con-

tributions of the remaining DNA atoms are known for urea (34) and are estimated for the other solutes using information for chemically similar protein atoms.

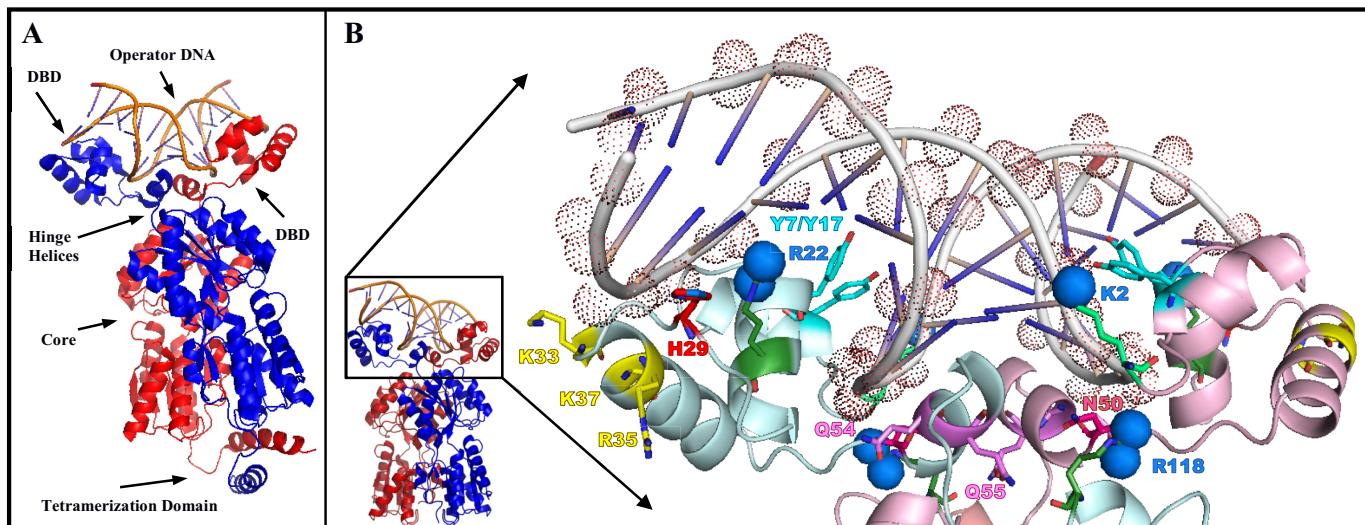
## Background on lac repressor–lac operator interaction

**R** is a homo-tetramer (dimer of dimers). Each subunit consists of a N-terminal 49-residue headpiece or DNA binding domain (**DBD**), a 13-residue flexible connector (hinge) region, a core regulatory domain and a C-terminal tetramerization domain (7,8,37). **DBDs** of two subunits of **R** bind one palindromic **O** sequence. The structure of **RO** in Figure 1 shows that these **DBD** interact with the **O** major groove and that the  $\alpha$ -helices formed by coupled folding of the hinge regions interact with the central minor groove of **O**, bending it away from the body of **R**. The entire assembly of **DBD** and hinge helices on **O** DNA binds to the core of **R**. This core-hinge helix-**DBD** interaction facilitates transmission to the **DBD** of the conformational change that occurs upon binding of allolactose or another inducer to the core regions of the **R** subunits, weakening **O** binding (14–16,38).

Interpretation of heat capacity and entropy changes in binding of lac **R** to **O** and **O** variants, as compared with binding to non-operator DNA, revealed extensive coupled folding to form the specific **RO** complex but not the non-operator (**RD**) complex (12,13), and NMR studies identified the two hinge regions as the residues that fold to form  $\alpha$ -helices in specific but not non-specific binding (9–11). Analysis of the strong destabilizing effect of urea and strong stabilizing effect of GB on the equilibrium constant  $K_{\text{obs}}$  for **RO** complex formation provided additional evidence for coupled folding ( $\alpha$ -helix formation with burial of amide oxygens) and dehydration of DNA phosphates (39,40).

**Key functional groups in R-R and R-O interfaces formed in the RO complex.** Analysis of the structure of **RO** reveals that aromatic, amide and cationic side chains of the **DBD** and hinge regions, together with amide backbone groups and non-polar (hydrophobic) side chains, are centrally involved in forming the new **R-R** and **R-O** interfaces in **RO** (7–11,41,42). In each half-site, two tyrosine residues (Y17, Y7) on the surface of the **DBD** interact with DNA bases in the major groove of **O**. Positively-charged groups (e.g. K2, R22, N-terminal –NH<sub>3</sub><sup>+</sup>) on each **DBD** appear positioned to interact favorably with (i.e. within 6 Å of) negatively charged DNA phosphate oxygens, as is R118 of the core repressor (43).

The hinge regions (residues 50–62) are disordered in unbound **R** but fold into a pair of interacting  $\alpha$ -helices in the process of **RO** formation. These insert in the central minor groove of **O**, burying side-chain amide groups of N-terminal residues N50 and Q54 in the minor groove interface and the Q55 amide in an interface with the other hinge region (backbone amide of V52) and the core (R118). Non-polar side chains of residues A53, L56, A57 contact sugars of bases 10–13 in minor groove, and residues V52, A53, Q55, L56 form apolar contacts where hinge helices pack against each other. Interactions of the hinge helices with the minor groove bends **O** DNA away from repressor by ~40° (7–11,41,42). Analysis of the effect of glycine betaine on the **RO** equilibrium constant revealed that anionic oper-



**Figure 1.** Key functional group determinants of *RO* Stability, Specificity and [Solute] Effects. (A) Structure of the specific complex between a pair of subunits of lac repressor and 16 bp lac *O* DNA (from PDB ID: 1Z04 (37)). (B) Key functional groups of the *RO* complex (PDB ID: 1EFA (7)). Aromatic rings of Y7 and Y17 (cyan) on the surface of each **DBD** of *R* are buried in the **DBD**-major groove interfaces of *RO*. Side-chain amide groups of N50 (dark pink) and Q55 (light pink), water-accessible in the flexible hinge regions of *R*, are buried in hinge helix and minor groove interactions in *RO*. Cationic N groups (blue spheres) of R22 (**DBD**) and R118 (both dark green) and K2 (light green) residues are within 6 Å of a DNA phosphate (red stippled spheres) in *RO*. In addition, cationic groups of K33, R35, K37 (all yellow) on the outer edge of the **DBDs** and also H29 (red) cationic groups are near DNA phosphate in *RO*. All these groups may have significant favorable coulombic interactions with DNA phosphates in *RO* (43).

ator phosphate oxygens are significantly dehydrated in the specific *RO* interface (39).

The mechanism of the *RO* interaction can be deduced from information about which of these key functional groups are buried in forming *TS* from *R* and *O* (*R* + *O* → *TS*), and which ones are buried subsequently in conversion of *TS* to *RO* (*TS* → *RO*).

## MATERIALS AND METHODS

### Chemicals and buffers

Urea, proline, glycerol, glycine betaine (GB) and tetraethylene glycol (tetraEG or TEG), purchased from Fluka and Sigma Aldrich were all of the highest available purity (>98%). Solutions of these solutes were prepared in the standard binding buffer (BB), which is 10 mM K<sub>2</sub>HPO<sub>4</sub>, 0.1 mM ethylenediaminetetraacetic acid (K<sub>2</sub>), 1 mM dithiothreitol (DTT), 100 µg/ml bovine serum albumin, 5% glycerol. KCl was added to obtain a final [K<sup>+</sup>] of 220 mM (for some tetraEG experiments) or 260 mM (for urea, proline, glycerol, GB and other tetraEG experiments), and pH was adjusted with HCl to pH 7.5 at assay temperature (25°C). The molar concentration of K<sup>+</sup> was held constant in experiments as a function of solute concentration.

### Lac repressor and lac operator DNA

Lac repressor tetramer was prepared, stored and diluted for use as previously described (40). Synthetic SymL DNA top strand (bases 1–36) and bottom strand (bases 1–40) were purchased in purified form from Integrated DNA Technologies (IDT) and annealed as described previously (13). The overhang was filled in by DNA polymerase Klenow fragment (44) using α-P<sup>32</sup> dATP and cold dCTP, followed

by a 2 µM cold dATP chase. The resulting double-stranded 40 bp duplex was purified on a 2 ml Sephadex G25 column run in TE buffer, and DNA concentration was determined in a micro cell using a Beckman DU640 spectrophotometer. The symmetric lac operator sequences are as follows.

O Sym36 Top  
5'-CACATATACC AATTGTGAGC GCTCACAAATT  
CCAGAT

O Sym40 Bottom  
5'-GTGTATCTGG AATTGTGAGC GCTCACAAATT  
GGTATATGTG

### Nitrocellulose filter binding assays

Filter binding assays were used to determine the amount of operator DNA retained on a nitrocellulose filter membrane in a complex with lac repressor protein (30,32). All thermodynamic and kinetic assays were performed in BB. For all assays, repressor was in sufficient excess over operator so only 1:1 *RO* complexes were formed. The fraction of operator DNA in the form of *RO* complexes (θ) was determined from the radioactive counts per minute (cpm) of <sup>32</sup>P-labeled operator DNA retained on the filter using the relationship

$$\theta = \frac{cpm - cpm_{bkgd}}{E(cpm_{tot} - cpm_{bkgd})} \quad (1)$$

where cpm<sub>bkgd</sub> is the cpm of <sup>32</sup>P-labeled operator DNA retained on the filter in the absence of lac *R*, cpm<sub>tot</sub> is the total cpm of <sup>32</sup>P-labeled operator DNA filtered in each assay and *E* is the filter efficiency (the fraction of *RO* complexes in solution that are retained on the filter membrane), determined as previously described (30). Average filter efficiency was 0.61 ± 0.09. For the range of solute concentrations studied here, we observe no systematic variation of filter efficiency

with solute concentration within the  $\pm 15\%$  uncertainty in these determinations.

To test if the low concentration of  $\text{RO}_2$  could contribute significantly to  $r$ -values, we analyzed TEG equilibrium titration data (Supplementary Figure S1), where the  $[\text{RJ}]:[\text{O}]$  total concentration ratio ( $\sim 2:1$ ) is smaller than in our other thermodynamic and kinetic studies. These titration data were fitted including and excluding 1:2 complexes and  $r$ -values were compared. The change in  $r$ -value from inclusion of 1:2 complexes is  $<5\%$ , which is negligible given the experimental uncertainty. For all other conditions, the  $[\text{RJ}]:[\text{O}]$  total concentration ratio is larger, and the effect will be even smaller.

Binding activity assays were performed to convert total repressor concentrations to active concentrations (44); 40–45% of repressor tetramers were active in binding operator.

**Equilibrium titrations with solutes.** For solute equilibrium titrations, Lac  $\text{R}$  and Lac  $\text{O}$  concentrations were selected to give initial fractional occupancies  $\theta$  (Equation 1) in the range  $0.22 < \theta < 0.38$  for stabilizing solutes, and  $\theta = 0.79 - 0.83$  for the destabilizing solute urea. The Lac  $\text{O}$  concentration ( $[\text{O}]_{\text{total}}$ ) was 20 pM except for experiments with tetraEG, where 10 pM was used. Total concentrations of active Lac  $\text{R}$  ( $[\text{R}_{\text{site,active}}]_{\text{total}}$ ) expressed on a dimer (operator binding site) scale, were usually in the range 68–85 pM, except for urea (1.7 nM). Series of samples with the same total repressor and operator concentrations and increasing solute concentration in the range 0 to 1 molar (1.2 molar for tetraEG) were prepared, incubated for 1–3 h and filtered. Values of  $\theta$  as a function of solute concentration (see Supplementary Figure S1A) were determined (Equation 1) and used to calculate equilibrium constants  $K_D$  as a function of solute molarity from Equation 2:

$$K_D = \left( (1 - \theta) \left( \frac{[\text{R}_{\text{site,active}}]_{\text{total}}}{\theta} - [\text{O}]_{\text{total}} \right) \right)_{eq} \quad (2)$$

Values of  $\ln K_D$  were fitted linearly with fixed intercepts, as functions of solute molarity. The intercepts ( $\ln K_D^0$ ; Supplementary Table S1) obtained from data in the absence of solutes, were used to obtain normalized plots of  $\ln K_D / \ln K_D^0$  versus solute concentration (Figure 2A). Propagated uncertainties were calculated for each data point and representative uncertainties are shown in figures.

**Decay to equilibrium kinetic studies.** Kinetic experiments were performed as a function of solute concentration by mixing 10–20 pM  $^{32}\text{P}$ -labeled operator DNA with excess repressor ( $[\text{R}_{\text{site,active}}]_{\text{total}}$ ) in the range 80–406 pM for all solutes. Repressor concentrations were chosen so that the binding reaction decayed to equilibrium ( $\theta_{eq} \leq 0.9$ ). Single-exponential decay-to-equilibrium kinetics (examples of which are shown in Supplementary Figure S1B) were analyzed using background corrected cpm (i.e. cpm–cpm<sub>bkgd</sub>) and the Equations (3,4).

$$\text{cpm}_t = \text{cpm}_{eq} (1 - e^{-k_{obs} t}) \quad (3)$$

where,

$$k_{obs} = k_a \left( [\text{R}_{\text{site,active}}]_{eq} + [\text{O}]_{eq} \right) + k_d \approx k_d \left( 1 + \frac{[\text{R}_{\text{site,active}}]_{\text{total}}}{K_D} \right) \quad (4)$$

Here  $[\text{R}_{\text{site,active}}]_{eq} = [\text{R}_{\text{site,active}}]_{\text{total}} - \theta_{eq}[\text{O}]_{\text{total}}$  and  $[\text{O}]_{eq} = (1 - \theta_{eq})[\text{O}]_{\text{total}}$ , so that  $([\text{R}_{\text{site,active}}]_{eq} + [\text{O}]_{eq}) = [\text{R}_{\text{site,active}}]_{\text{total}} + [\text{O}]_{\text{total}} (1 - 2\theta_{eq})$ . Since  $[\text{O}]_{\text{total}} \ll [\text{R}_{\text{site,active}}]_{\text{total}}$ , the term  $([\text{O}]_{\text{total}} (1 - 2\theta_{eq}))$  can be neglected. Dissociation rate constants  $k_d$  were obtained from decay to equilibrium rate constants  $k_{obs}$  using the appropriate  $[\text{R}_{\text{site,active}}]_{\text{total}}$  and the  $K_D$  obtained from the corresponding solute titration, and association rate constants  $k_a$  were obtained from  $k_a = k_d/K_D$ . Values of  $k_d$  and  $k_a$  obtained from tetraEG experiments at low salt concentration (220 mM), were corrected for salt effects on the two rate constants using experimentally determined values of  $\text{Sk}_i$  (23). Normalized plots of  $\ln k_d$  and  $\ln k_a$  as functions of solute molarity (Figure 2B and C) were obtained by using  $\ln k_d^0$  and  $\ln k_a^0$  values obtained from experiments in absence of solutes (Supplementary Table S1). Contribution of the term  $([\text{O}]_{\text{total}} (1 - 2\theta_{eq}))$  in comparison to  $[\text{R}_{\text{site,active}}]_{\text{total}}$  was estimated from experimental concentrations and values of  $\theta$  and found to be on average 4%, which is negligible. Propagated uncertainties ( $\sigma$ ) calculated for each data point using Equation 5, are shown in figures.

$$\sigma_f^2(x, y, z) = \left( \frac{\partial f}{\partial x} \right)^2 \sigma_x^2 + \left( \frac{\partial f}{\partial y} \right)^2 \sigma_y^2 + \left( \frac{\partial f}{\partial z} \right)^2 \sigma_z^2 \quad (5)$$

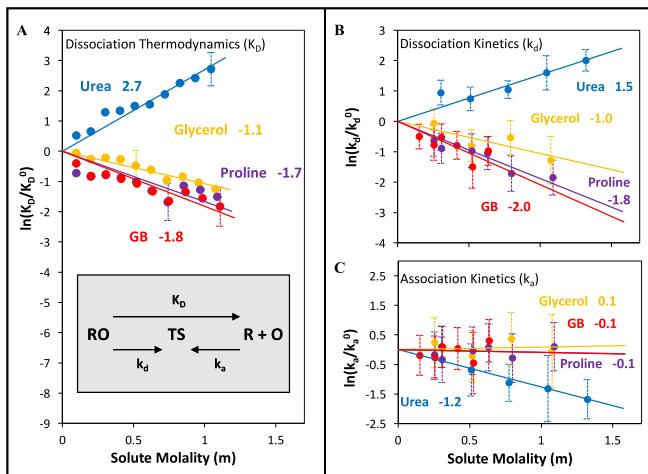
Equation (5) shows the generic method for calculating propagated uncertainty ( $\sigma_f$ ) for any given function,  $f(x, y, z)$  using the product of partial derivatives of the function ( $f$ ) with respect to the independent variables ( $x, y, z$ ) and the uncertainties associated with each variables ( $\sigma_x, \sigma_y, \sigma_z$ ).

## RESULTS

### Solute and salt effects on the $\text{RO}$ dissociation equilibrium constant ( $K_D$ )

Effects of probe solutes on the  $\text{RO}$  dissociation equilibrium constant  $K_D$  are shown in Figure 2A (and Supplementary Figure S2), in which the logarithm of  $K_D$ , obtained from solute titrations, is plotted versus molal solute concentration ( $m_3$ ). These plots are linear over the range studied (0.1 to  $\sim 1.5$  m). Effects of solutes are characterized by slopes  $d\ln K_D/dm_3$ , here called  $r_K$ -values (Figure 2A). These  $r_K$ -values, reported with uncertainties in Supplementary Table S2, range from 2.7 for urea to  $-1.1$  for glycerol,  $-1.7$  for proline and  $-1.8$  for GB. Data for a fourth stabilizing solute (tetraethylene glycol, TEG), obtained at a somewhat lower KCl concentration, are reported in Supplementary Figure S2. The TEG  $r_K$ -value ( $-2.0$ ) is similar to that of GB. These  $r_K$ -values are proportional to thermodynamic  $m$ -values ( $m$ -value =  $d\Delta G^\circ/dm_3 = -RTd\ln K_D/dm_3 = -RT(r_K\text{-value}) = \Delta\mu_{23}$ , where  $\Delta\mu_{23}$  is the difference in chemical potential-derivatives  $\mu_{23} = \partial\mu_2/\partial m_3$  at constant pressure, temperature and  $m_2$ , between the  $\text{RO}$  complex and reactants (28)). Values of  $\Delta\mu_{23}$ , and hence  $r_K$ -values, are interpreted as preferential interactions of the solute with the functional groups exposed in the process (28–32,45) (see Analysis below and Supplementary Equations S4 and 5).

Figure 2A shows that addition of urea favors  $\text{RO}$  dissociation, increasing  $K_D$  by 15-fold at 1 molal urea. This is expected because urea is found to interact favorably with all protein and nucleic acid functional groups except



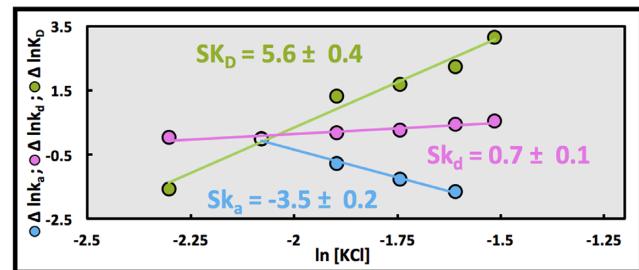
**Figure 2.** Solute effects on the equilibrium and rate constants of the **RO** interaction. Logarithms of equilibrium dissociation constants ( $K_D$ , Panel A) and dissociation and association rate constants ( $k_d$ , Panel B and  $k_a$ , Panel C), normalized by their zero-solute values as described in the ‘Materials and Methods’ section, are plotted versus molal solute concentration for urea, glycerol, proline and glycine betaine. Linear fits and  $r$ -values (semi-log slopes  $d\ln K_D/dm_3$  (A);  $d\ln k_d/dm_3$  (B);  $d\ln k_a/dm_3$  (C)) are shown (see Supplementary Table S2). Uncertainties represent one standard deviation determined as in the ‘Materials and Methods’ section.

cationic N (Supplementary Table S3) (33,34,46). Other solutes (glycerol, proline, GB, TEG) all stabilize **RO**, reducing  $K_D$  by 3- to 7-fold at 1 molal. Effects of urea and GB are similar to those obtained previously by different methods and under different conditions (28,39,40). Unlike urea, the **RO**-stabilizing solutes all exhibit wide ranges of favorable and unfavorable interactions with protein and nucleic acid atoms (Supplementary Table S3). All interact strongly and unfavorably with DNA anionic phosphate oxygens and protein amide and carboxylate oxygens, but interact favorably with aromatic C and amide N (33–36). All except interior repeat groups of TEG interact unfavorably with aliphatic ( $sp^3$ ) C (36). Unfavorable interactions of these solutes with oxygens and also (except for TEG) with hydrocarbon groups exposed in **RO** dissociation must be dominant over favorable interactions with aromatic C and amide N to give rise to their significant **RO**-stabilizing effects.

A log–log plot of the dissociation equilibrium constant  $K_D$  as a function of KCl concentration, obtained by Whitson and Matthews (23) using 40 bp **O** DNA, is shown in Figure 3. The slope  $SK_D = d\ln K_D/d\ln[KCl] = 5.6$  (Supplementary Table S4). The log–log functional form and the large destabilizing effect of [KCl] on the **RO** complex are the expected polyelectrolyte effect; Coulombic interactions between cationic groups on **R** and negatively charged phosphates of the **O** polyanion are disfavored at higher [KCl].

#### Solute and salt effects on repressor–operator association and dissociation rate constants

Solute effects on the kinetics of association of **R** and **O** and dissociation of **RO** are quantified by kinetic  $r_k$ -values  $d\ln k_i/dm_3 = -(1/RT) d\Delta G^{\ddagger}/dm_3$  where  $k_i$  is either the association ( $k_a$ ) or dissociation ( $k_d$ ) rate constant. Dissociation rate constants ( $k_d$ ) were determined from decay-to-



**Figure 3.** Comparison of salt (KCl) effects on the equilibrium and rate constants of the **RO** interaction. Logarithms of equilibrium dissociation constant ( $K_D$ ) and association and dissociation rate constants ( $k_a$ ,  $k_d$ ) determined by Whitson and Matthews (23) are plotted versus the logarithm of the molar KCl concentration. All  $K_D$ ,  $k_a$  and  $k_d$  values are normalized relative to best fit values of these quantities at 0.125 M salt. Log–log slopes  $SK_D$ ,  $Sk_a$  and  $Sk_d$  and uncertainties are shown (see Supplementary Table S4).

equilibrium experiments as described in the ‘Materials and Methods’ section, and association rate constants ( $k_a$ ) were determined from dissociation rate constants and equilibrium constants at each solute concentration.

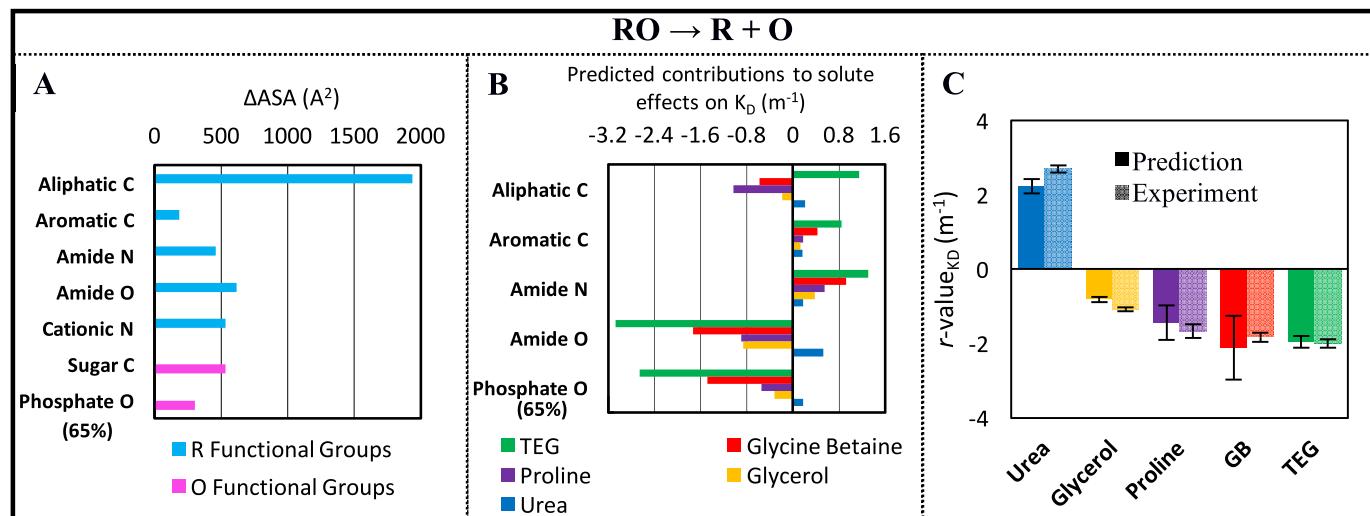
Plots of  $\ln k_d$  and  $\ln k_a$  versus solute molality are shown in Figure 2B and C, respectively. These plots are linear in the range of solute molality investigated (0.1 to ~1.5 m). Molal scale  $r_k$ -values for all solutes are indicated in Figure 2B and C and reported with uncertainties in Supplementary Table S2. The association kinetic  $r_k$ -value is interpreted as the effect of the solute on the process of converting reactants (**R**, **O**) to **TS**: the dissociation kinetic  $r_k$ -value is interpreted as the effect of the solute on the process of converting **RO** to this same **TS**.

Figure 2B reveals that dissociation kinetic  $r_k$ -values exhibit the same pattern as thermodynamic  $r_K$ -values (Figure 2A). The **RO**-destabilizing solute urea increases the dissociation rate constant  $k_d$  while the **RO**-stabilizing solutes reduce  $k_d$ . Indeed, dissociation kinetic  $r_k$ -values for the **RO**-stabilizing solutes are the same within uncertainty as thermodynamic  $r_K$ -values, while the urea  $r_k$ -value is about half as large as the urea  $r_K$ -value.

These findings indicate that the entire effect of the **RO**-stabilizing solutes is on the conversion of the high free-energy **TS** to **RO**, while approximately half the destabilizing effect of urea is on conversion of **TS** to **RO**. Consistent with this, Figure 2C reveals that association kinetic  $r_k$ -values for the **RO**-stabilizing solutes are zero within uncertainty, while the urea association kinetic  $r_k$ -value is comparable in magnitude to its dissociation kinetic  $r_k$ -value.

Literature values of the rate constants  $k_a$  and  $k_d$  as functions of [KCl] for the **RO** interaction involving 40 bp **O** are shown in Figure 3 as log–log plots (23). At 0.15 M salt, the association rate constant ( $\sim 10^9 \text{ M}^{-1}\text{s}^{-1}$  or larger) and activation energy of association ( $\sim 4 \text{ kcal}$  (23)) are consistent with a diffusion-limited reaction. An increase in [KCl] destabilizes **RO** by reducing the association rate constant greatly ( $Sk_a = -3.5$ ) and increasing the dissociation rate constant slightly ( $Sk_d = 0.7$ ) (Supplementary Table S4).

At 0.26 M KCl, the salt concentration investigated here, we find  $k_a = 4 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$  in the absence of perturbing solutes, two orders of magnitude less than the diffusion limit



**Figure 4.** Contributions of Different **R** and **O** Atom types to solute effects on **RO** dissociation. (A) Composition of the **R** (blue) and **O** (pink) surfaces exposed in **RO** dissociation ( $\Delta\text{ASA}_{\text{total}} = 5800\text{\AA}^2$ ). (B) Predicted most significant contributions (in  $\text{m}^{-1}$ ) to observed solute effects ( $r_K$ -value =  $d\ln K_D/dm_3$ ) in **RO** dissociation. For **O** anionic phosphate oxygen, predictions are shown for a corrected (scaled)  $\Delta\text{ASA}$  to correct for incomplete dehydration ( $302\text{\AA}^2$ ; see text). (C) Comparison of experimental and predicted overall effect of solutes on equilibrium constant for **RO** dissociation ( $K_D$ ).

(Supplementary Table S1). Because most of the KCl dependence of  $K_D$  is in  $k_a$  and not in  $k_d$ , therefore most of the favorable Coulombic interactions between cationic groups in lac **R** and anionic phosphates in lac **O** occur in initial binding of **R** to **O** and subsequent conformational changes to form the high-free-energy transition state ( $\mathbf{R} + \mathbf{O} \rightarrow \mathbf{TS}$ ) (47).

## DISCUSSION

### Analysis of thermodynamic $r_K$ -values for **RO** dissociation using $\Delta\text{ASA}$ information and solute $\alpha$ -values

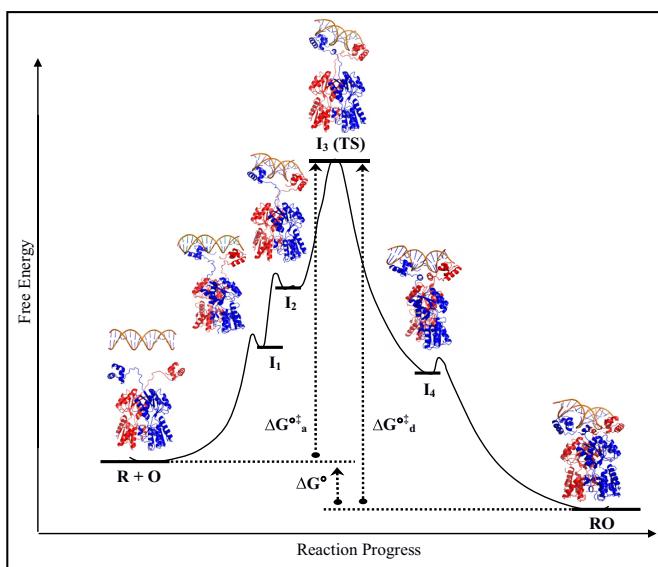
Thermodynamic ( $K_D$ )  $r_K$ -values quantifying effects of solutes on dissociation of **RO** are determined by preferential interactions of these solutes with the different types of unified atoms on the **R** and **O** surfaces exposed in dissociation. Using the **RO** structure (1EFA (7)) and models for free **R** and **O** (models described in Supplementary Methods), the amount and composition of **R** and **O** surfaces exposed in **RO** dissociation were estimated. The overall change in ASA ( $\Delta\text{ASA}$ ) is predicted to be approximately  $5800\text{\AA}^2$  (Figure 4A). The contribution to  $\Delta\text{ASA}$  from **R** in dissociation ( $4100\text{\AA}^2$ ) greatly exceeds that from **O** ( $1700\text{\AA}^2$ ). Exposure of  $3000\text{\AA}^2$  of aliphatic ( $\text{sp}^3$ ) and aromatic ( $\text{sp}^2$ ) C (70% from **R**, 30% from **O**) accounts for half of the overall  $\Delta\text{ASA}$  of **RO** dissociation. The predicted contribution of anionic DNA phosphate oxygens to the  $\Delta\text{ASA}$  is  $467\text{\AA}^2$ , one-fourth of the DNA contribution to  $\Delta\text{ASA}$ . Other significant contributions to the  $\Delta\text{ASA}$  come from **R** amide O, N and cationic N atoms and **O** sugar  $\text{sp}^3$  C.

Predicted contributions to the observed solute  $r_K$ -values from solute interactions with all types of **R** and **O** atoms exposed in dissociation of **RO** are shown in the bar graph of Figure 4B. These are calculated from the ASA information of Figure 4A and the interaction potentials ( $\alpha$ -values) of Supplementary Table S3. These  $\alpha$ -values quantify solute interactions with a unit area of the seven major types of uni-

fied atoms of protein (aliphatic and aromatic C; amide and cationic N; amide, hydroxyl and carboxylate O) and with anionic phosphate O of DNA (28,33–36). Supplementary Table S3 also lists  $\alpha$ -values for interaction of urea with other types of DNA atoms.  $\alpha$ -Values for interactions of other solutes with these types of DNA atoms are assumed to be the same as those for chemically-similar types of protein atoms. Inspection of urea  $\alpha$ -values shows this is a reasonable first approximation.

Figure 4 reveals that while aliphatic carbon (C) makes the largest contribution to  $\Delta\text{ASA}$  (Figure 4A), exposure of DNA anionic phosphate oxygens in **RO** dissociation is predicted to make the largest contribution to the overall  $r_K$ -value for the three solutes (TEG, GB and proline) that most stabilize **RO** (Figure 4B), because of the large positive  $\alpha$ -values and consequent strong unfavorable interactions of these solutes with phosphate oxygen. For GB and TEG, the two solutes that are most highly excluded from anionic phosphate oxygens, predicted contributions to  $r_K$ -values from anionic phosphate oxygen  $\Delta\text{ASA}$  appear too large in magnitude, causing magnitudes of overall predicted  $r_K$ -values to exceed observed  $r_K$ -values by 60 to 70% (Supplementary Table S5). A likely interpretation of these discrepancies is that anionic phosphate oxygens are not completely dehydrated when they are incorporated into the **RO** interface, retaining  $\sim 1/3$  of their original water of hydration. Hence the amount of water that is taken up upon exposure of these phosphate oxygens in **RO** dissociation is only  $2/3$  of that calculated from the structure. The thermodynamically-detected residual hydration of phosphates in the interface is most simply parameterized for analysis by reducing the anionic phosphate oxygen  $\Delta\text{ASA}$  for **RO** dissociation by approximately one-third, from  $467\text{\AA}^2$  to  $302 \pm 32\text{\AA}^2$ .

Structural evidence for residual water in the interface of a complex between one **DBD** and a half operator site was obtained previously (41), though more recent structures of the **RO** complex formed by two **DBD** do not resolve any bound



**Figure 5.** Free energy versus progress schematic with proposed intermediates and **TS** in **RO** formation. Proposed sequence of intermediates (**I<sub>1</sub>**–**I<sub>4</sub>**) in the structure-based mechanism of formation/dissociation of the **RO** complex (see text) are shown on a schematic activation diagram of free energy versus progress. The observed single-exponential kinetics of association and dissociation justify showing all intermediates as unstable and in rapid equilibrium with earlier intermediates and reactants. Intermediates are placed on the progress axis in order of their  $\Delta\text{ASA}$  values; actual free energy differences between these species are unknown. **I<sub>3</sub>**, in which nucleation but not propagation of the hinge helices has occurred, is deduced to be the closest of the **I**-species to the high-free-energy transition state **TS**, on the basis of the analysis presented in the text.

water. Hydration of anionic phosphate oxygens is estimated to be at least three layers of water (40), so a residual hydration of one layer of water at anionic phosphate O would explain the discrepancy between predicted and observed  $r_K$ -values for GB and TEG. Thermodynamic and structural evidence has been presented for water in the interface of other protein–nucleic acid complexes (41,48,49). This bound water is proposed to contribute significantly (about 18 cal/K per mole for each bound water) to the large positive heat capacity change of dissociation (50). Monolayer coverage of the 467 Å<sup>2</sup> of DNA phosphate oxygen ASA in the interface of the **RO** complex would require that ~47 water molecules remain bound, proposed to contribute approximately 0.8 kcal/K to the heat capacity change of **RO** dissociation. This value is comparable to the amount by which the observed heat capacity change for **RO** dissociation (1.5 ± 0.2 kcal/K) (39) exceeds that predicted from the  $\Delta\text{ASA}$  of hydrocarbon and polar groups (0.7 kcal/K).

Supplementary Table S5 compares observed solute  $r_K$ -values with those predicted using the parameterized contribution of anionic phosphate oxygen (Figure 4C); this phosphate correction greatly improves agreement between predicted and observed  $r_K$ -values for GB and TEG without degrading it significantly for the other three solutes (Supplementary Figure S3A and B).

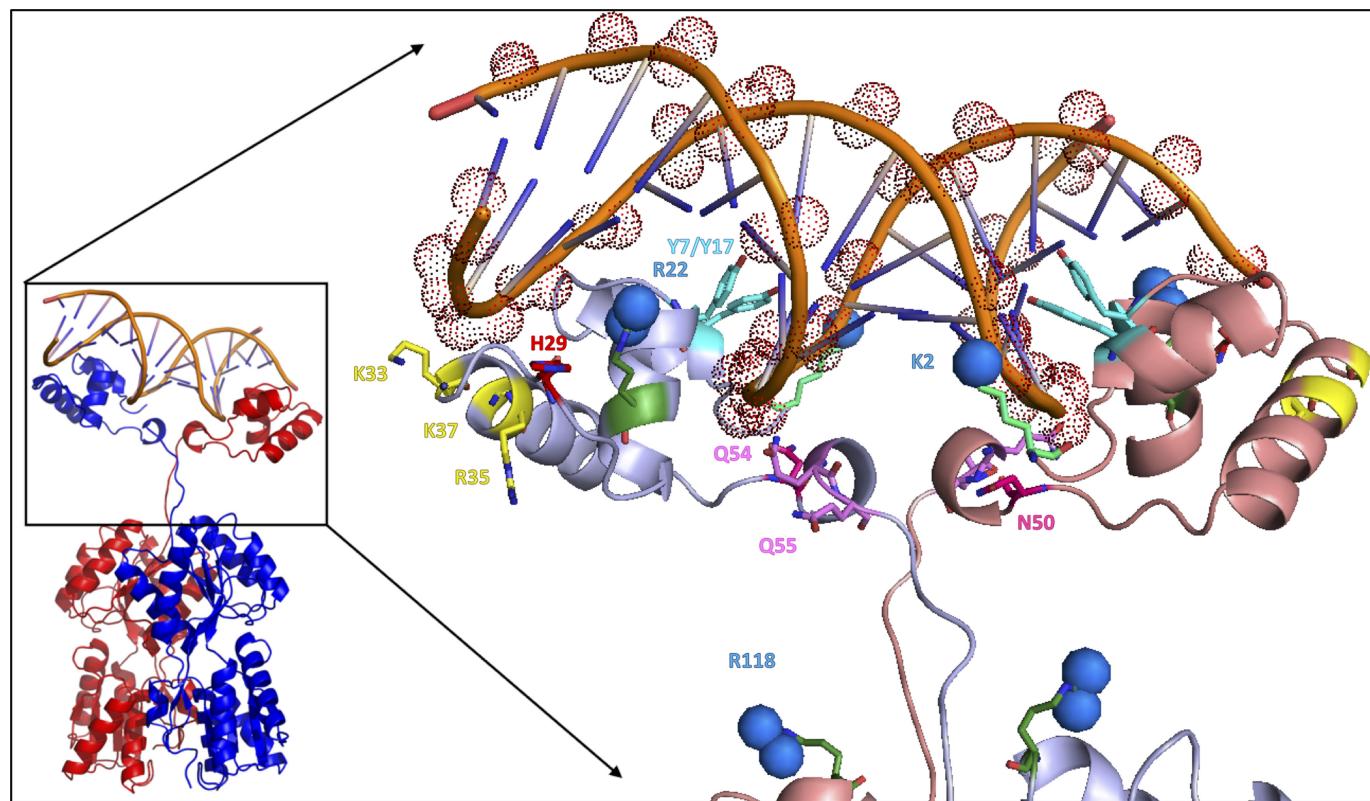
### NLPB prediction of $SK_D$ quantifying the KCl effect on $K_D$ of **RO** dissociation

Coulombic free energies  $G_{\text{coul}}^o$  of **R**, **O** and the **RO** complex at 25°C were obtained as functions of 1:1 salt (KCl) concentration from the non-linear Poisson-Boltzmann equation as described in Supplementary Data. These energies  $G_{\text{coul}}^o$  are plotted versus  $\ln[\text{KCl}]$  in Supplementary Figure S4A. The corresponding Coulombic free energy changes in **RO** dissociation,  $\Delta G_{\text{coul}}^o = G_{\text{coul},R}^o + G_{\text{coul},O}^o - G_{\text{coul},\text{RO}}^o$  are plotted versus  $\ln[\text{KCl}]$  in Supplementary Figure S4B. Individual values of free energies  $G_{\text{coul}}^o$  decrease non-linearly with increasing  $\ln[\text{KCl}]$ , but  $\Delta G_{\text{coul}}^o$  decreases linearly with  $\ln[\text{KCl}]$  in the range of interest.  $\frac{d\Delta G_{\text{coul}}^o}{d\ln[\text{KCl}]} = -RT SK_D$  yields a parameter-free predicted value of  $SK_D = 6.5$ , which agrees well with the experimental  $SK_D = 5.6 \pm 0.4$  (Figure 3; Supplementary Figure S3C and Table S4; ref (23)).

### Qualitative interpretation of effects of KCl and solutes on rate constants

At 0.26 M KCl, the second order association rate constant  $k_a$  for **RO** formation is approximately two orders of magnitude less than the diffusion limit. From Figure 2B and C, the effect of urea concentration on the dissociation equilibrium constant  $K_D$  is distributed roughly equally between the rate constants  $k_a$  and  $k_d$ . From Figure 3, most of the large effect of KCl concentration on  $K_D$  is on  $k_a$ ;  $k_d$  is relatively independent of KCl concentration. Taken together, these findings indicate that at 0.26 M KCl the rate determining step in **RO** formation is subsequent to the initial diffusion–collision interaction of **R** and **O**. This initial binding step is deduced to be in rapid equilibrium on the time scale of subsequent conformation changes, and the high-free-energy transition state for the rate-determining conformational change is somewhere in mid-mechanism of **RO** formation.

Effects of the three **RO**-stabilizing solutes on the kinetics are largely on the dissociation rate constant  $k_d$ . Comparison of Figure 2B and C with Figure 2A shows that these solutes have very similar effects on  $k_d$  and  $K_D$  and have negligibly small effects on  $k_a$ . The most likely interpretation of the small-magnitude  $k_a$   $r$ -values for **RO**-stabilizing solutes is that (i) little if any dehydration of anionic phosphate oxygens occurs in forming **TS** from **R** and **O**, and (ii) there is near-complete compensation between **TS**-stabilizing and **TS**-destabilizing interactions of these **RO**-stabilizing solutes. If  $k_a$  were closer to the diffusion limit and if KCl and urea also had small effects on  $k_a$  and large effects on  $k_d$ , an alternative interpretation would be that **TS** is very early, involving few interactions between **R** and **O** charges or functional groups and no dehydration of interfaces. But the KCl and urea results, and the fact that  $k_a$  is much less than the diffusion limit, show that this interpretation is incorrect. The only atoms that these **RO**-stabilizing solutes interact favorably with are aromatic carbon ( $sp^2$  C) and amide nitrogen (N). We therefore assume that anionic phosphate oxygens remain hydrated in **TS**, and propose that the most likely **TS** candidates are ones that preferentially bury  $sp^2$  C and amide N, but not amide O.



**Figure 6.** Structural model of the high free energy *TS* in *RO* binding/dissociation. Major functional groups (protein and DNA) deduced to be involved in forming *TS* are displayed. Aromatic C groups from Y7 and Y17 (cyan) are buried in forming the **DBD**-DNA interface. Amide N groups from N50 (dark pink), Q54 and Q55 (light pink) residues are buried in nucleation of hinge helix. Cationic N groups (blue spheres) from R22 (dark green) and K2 (light green) residues in repressor protein as well as other potential charge contributing groups from lysine and arginine (K33, R35, K37; yellow) on the outer edge of the **DBDs**, and cationic N from a histidine residue (H29; red) in each **DBD** make contacts with phosphate oxygen groups (red stippled spheres) on operator DNA backbone. The cationic N (blue spheres) of R118 on the core of the protein are not close enough to operator DNA phosphates to contribute significantly to the salt dependence of *TS* formation from *R* and *O*.

#### Evidence for the nature of the high-free-energy transition state from comparison of predicted and observed solute and salt effects on rate constants $k_a$ , $k_d$

Structural models of four possible intermediates in *RO* complex formation from free *R* and *O* (labeled *I*<sub>1</sub> to *I*<sub>4</sub>; see Figure 5) were generated starting from published coordinates of non-specific (*RD*) and specific (*RO*) complexes (see Supplementary Methods). These models correspond to steps in the proposed mechanism (see ‘Introduction’ section), deduced from structural information. Intermediate *I*<sub>1</sub> is based on the structure of the non-specific (*RD*) complex (10) (using the *ODNA* sequence), which we assume is an appropriate model of the least advanced intermediate formed with *O* DNA. Modeled intermediate *I*<sub>2</sub> is more advanced than *I*<sub>1</sub>, having the specific interactions of the two *R* **DBBs** with the major grooves of *O* that are present in the stable *RO* complex. In addition to the **DBB**-major groove interactions of *I*<sub>2</sub>, intermediate *I*<sub>3</sub> has nucleated  $\alpha$ -helix formation at the N-termini of the two hinge regions, contacting and bending the central minor groove of *O*. The hinge helices are completely formed but not yet inserted in the bent *O* minor groove in more advanced intermediate *I*<sub>4</sub>, in which the **DBB**-hinge helix-*O* DNA module is also not yet docked on the core of repressor to form the stable *RO* complex.

Solute  $r_K$ -values for formation of each intermediate from free *R* and *O* and from *RO* are predicted from the  $\alpha$ -values of Supplementary Table S3 and the results of ASA analysis of these models of possible intermediates (Supplementary Table S6). Comparison of these predictions (Supplementary Figure S5) with observed association and dissociation kinetic  $r_k$ -values and with the  $r_K$ -value of *RO* reveals when DNA phosphates lose their outer layers of hydration in *RO* formation and which, if any, of these proposed intermediates might be a good model for the high-free-energy transition state that determines  $k_a$  and  $k_d$ .

*When in RO formation do DNA phosphates lose their outer layers of hydration?* Dehydration of the outer layers of hydration of DNA anionic phosphate oxygens (from three layers of water to one layer) must occur post-*TS* in *RO* formation. Conversely, hydration of anionic phosphate oxygens from one layer to three layers of water occurs pre-*TS* in *RO* dissociation. If substantial dehydration of anionic phosphate oxygen surface occurred pre-*TS* in association, *RO*-stabilizing solutes would of necessity also stabilize *TS* and therefore increase  $k_a$ . For *RO*-stabilizing solutes to have the observed minimal effect on  $k_a$ , the anionic phosphate oxygens that are involved in coulombic interactions with cationic groups of *R* must remain hydrated in associa-

tion to **TS**. Loss of two layers of hydration water of anionic phosphate oxygens in converting **TS** to **RO** makes a major (dominant) contribution to the  $k_d r_k$ -values, and therefore is a large part of the explanation for why  $k_d r_k$ -values and  $K_d r_K$ -values are so similar for these **RO**-stabilizing solutes. Conversion of **TS** to **RO** therefore tightens the interfaces between **R** domains and **O** phosphates, squeezing out two of the three layers of water of hydration of anionic phosphate oxygens.

*Nucleation of hinge helix formation in **I**<sub>3</sub> makes it the most plausible candidate for **TS**.* For the range of possible **TS** models considered here (**I** species **I**<sub>1</sub> to **I**<sub>4</sub> in Figure 5), Supplementary Figure S6 predicts contributions to the association kinetic  $r_k$ -value from the various atom-types buried in forming these **TS** candidates from **R** and **O** and compares predicted and observed association kinetic  $r_k$ -values. In all cases, anionic phosphate oxygens are assumed to remain completely hydrated in **TS** (0% dehydration). For comparison, a ‘late **I**’ model for **TS** is also considered, with the structure of **RO** but with 0% dehydration of anionic phosphate oxygens. The circled results in Supplementary Figure S6, representing the best agreement between predicted and observed association kinetic  $r_k$ -values, indicate that the most likely **TS** structure is more advanced than **I**<sub>1</sub> but less advanced than **I**<sub>4</sub>, and could be either **I**<sub>2</sub> or **I**<sub>3</sub>. The urea data are most definitive, favoring **I**<sub>3</sub> as **TS** though not excluding **I**<sub>2</sub>. Results for the **RO**-stabilizing solutes in Supplementary Figure S6 show a small but significant preference for **I**<sub>3</sub> or **I**<sub>2</sub> as **TS** over **I**<sub>1</sub> and **I**<sub>4</sub>.

Nucleation of hinge helix formation in **I**<sub>3</sub> makes it an attractive candidate for **TS** in **RO** formation. Nucleation steps generally are highly unfavorable. Nucleation of tertiary structure formation appears to be the high free energy **TS** in protein folding (29). Moreover, compensations between favorable and unfavorable interactions of **RO**-stabilizing solutes with the ΔASA of forming **I**<sub>3</sub> (also **I**<sub>2</sub>) from **R** and **O**, shown in Figure 6 and quantified in Supplementary Figure S6, result in small-magnitude predicted association kinetic  $r_k$ -values for these **RO**-stabilizing solutes, consistent with experiment.

Figure 6 shows our proposal that nucleation of hinge helix formation in the conversion of **I**<sub>2</sub> to **I**<sub>3</sub> buries side-chain amide N of three asparagine and glutamine residues (N50 (dark pink), Q54 and Q55 (light pink)) at the N-terminus of each hinge region, without much burial of amide O. Indeed, half of the amide N ASA buried in **RO** formation occurs in forming **I**<sub>3</sub>, our **TS** candidate, while only a third of total amide O burial occurs in forming **TS** (see Supplementary Table S6). Earlier in formation of this **TS** model from **R** and **O**, aromatic C atoms of tyrosines Y7 and Y17 (cyan) of each **DBD** are buried in the major grooves of **O** DNA (Figure 6), accounting for all the **R** aromatic C buried in the stable **RO** complex.

Post-**TS**, the favorable propagation steps of hinge helix formation occur. We propose that the pair of interacting hinge helices insert into the central minor groove of the **O** DNA, bending the DNA away from the protein and forming intermediate **I**<sub>4</sub>. In the final step, the assembly of **DBD**, **O** DNA and hinge-helices in **I**<sub>4</sub> docks on the C-terminal core of **R**, tightening and partially dehydrating the **R-O** in-

terface to form the final **RO** complex. Figure 5 places the various **I**-species and the proposed **TS** on an activation free energy diagram and summarizes this non-covalent self-assembly (nucleation-propagation) mechanism of **RO** formation.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Elliot Schad for his contributions to these solute kinetic experiments, and the reviewers for their careful reading and valuable comments on the manuscript.

## FUNDING

National Institutes of Health [GM118100, GM47022, GM103061]. Funding for open access charge: National Institutes of Health.

*Conflict of interest statement.* None declared.

## REFERENCES

- Muller-Hill,B. (1996) *The lac operon: a short history of a genetic paradigm*. Walter de Gruyter, Berlin.
- Jacob,F. and Monod,J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, **3**, 318–356.
- von Hippel,P.H., Revzin,A., Gross,C.A. and Wang,A.C. (1974) Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. *Proc. Nat. Acad. Sci. U.S.A.*, **71**, 4808–4812.
- O’Gorman,R.B., Rosenberg,J.M., Kallai,O.B., Dickerson,R.E., Itakura,K., Riggs,A.D. and Matthews,K.S. (1980) Equilibrium binding of inducer to lac repressor-operator DNA complex. *J. Biol. Chem.*, **255**, 10107–10114.
- Zwieb,C., Kim,J. and Adhya,S. (1989) DNA bending by negative regulatory proteins: Gal and Lac repressors. *Genes Dev.*, **3**, 606–611.
- Kercher,M.A., Lu,P. and Lewis,M. (1997) Lac repressor-operator complex. *Curr. Opin. Struct. Biol.*, **7**, 76–85.
- Bell,C.E. and Lewis,M. (2000) A closer view of the conformation of the lac repressor bound to operator. *Nat. Struct. Biol.*, **7**, 209–214.
- Lewis,M., Chang,G., Horton,N.C., Kercher,M.A., Pace,H.C., Schumacher,M.A., Brennan,R.G. and Lu,P. (1996) Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science*, **271**, 1247–1254.
- Spronk,C.A., Folkers,G.E., Noordman,A.M., Wechselberger,R., van den Brink,N., Boelens,R. and Kaptein,R. (1999) Hinge-helix formation and DNA bending in various lac repressor-operator complexes. *EMBO J.*, **18**, 6472–6480.
- Kalodimos,C.G., Biris,N., Bonvin,A.M., Levandoski,M.M., Guennuegues,M., Boelens,R. and Kaptein,R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science*, **305**, 386–389.
- Sliper,M., Bonvin,A.M., Boelens,R. and Kaptein,R. (1996) Refined structure of lac repressor headpiece (1-56) determined by relaxation matrix calculations from 2D and 3D NOE data: change of tertiary structure upon binding to the lac operator. *J. Mol. Biol.*, **259**, 761–773.
- Spolar,R.S. and Record,M.T. Jr (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
- Frank,D.E., Saecker,R.M., Bond,J.P., Capp,M.W., Tsodikov,O.V., Melcher,S.E., Levandoski,M.M. and Record,M.T. Jr (1997) Thermodynamics of the interactions of lac repressor with variants of the symmetric lac operator: effects of converting a consensus site to a non-specific site. *J. Mol. Biol.*, **267**, 1186–1206.
- Xu,J. and Matthews,K.S. (2009) Flexibility of inducer binding region is crucial for allostery in the *Escherichia coli* lactose repressor. *Biochemistry*, **48**, 4988–4998.

15. Swint-Kruse,L. and Matthews,K.S. (2009) Allostery in the LacI/GalR family: variations on a theme. *Curr. Opin. Microbiol.*, **12**, 129–137.
16. Butler,A.P., Revzin,A. and von Hippel,P.H (1977) Molecular parameters characterizing the interaction of *Escherichia coli* lac repressor with non-operator DNA and inducer. *Biochemistry*, **16**, 4757–4768.
17. Laiken,S.L., Gross,C.A. and von Hippel,P.H. (1972) Equilibrium and kinetic studies of *Escherichia coli* lac repressor-inducer interactions. *J. Mol. Biol.*, **66**, 143–155.
18. Riggs,A.D., Bourgeois,S. and Cohn,M. (1970) The lac repressor-operator interaction. 3. Kinetic studies. *J. Mol. Biol.*, **53**, 401–417.
19. Richter,P.H. and Eigen,M. (1974) Diffusion controlled reaction rates in spheroidal geometry. Applications to repressor-operator association and membrane bound enzymes. *Biophys. Chem.*, **2**, 255–263.
20. Winter,R.B., Berg,O.G. and von Hippel,P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor-operator interaction: kinetic measurements and conclusions. *Biochemistry*, **20**, 6961–6977.
21. Barkley,M.D. (1981) Salt dependence of kinetics of the lac repressor-operator interaction: role of nonoperator deoxyribonucleic acid in the association reaction. *Biochemistry*, **20**, 3833–3842.
22. von Hippel,P. H. and Berg,O.G. (1989) Facilitated target location in biological systems. *J. Biol. Chem.*, **264**, 675–678.
23. Whitson,P.A., Olson,J.S. and Matthews,K.S. (1986) Thermodynamic analysis of the lactose repressor-operator DNA interaction. *Biochemistry*, **25**, 3852–3858.
24. Falcon,C.M. and Matthews,K.S. (1999) Glycine insertion in the hinge region of lactose repressor protein alters DNA binding. *J. Biol. Chem.*, **274**, 30849–30857.
25. Falcon,C.M. and Matthews,K.S. (2000) Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. *Biochemistry*, **39**, 11074–11083.
26. Schramm,V.L. (2013) Transition states, analogues and drug development. *ACS Chem. Biol.*, **8**, 71–81.
27. Schramm,V.L. (2011) Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annu. Rev. Biochem.*, **80**, 703–732.
28. Record,M.T. Jr, Guinn,E., Pegram,L. and Capp,M. (2012) Introductory lecture: interpreting and predicting Hofmeister salt ion and solute effects on biopolymer and model processes using the solute partitioning model. *Faraday Disc.*, **160**, 9–44.
29. Guinn,E.J., Kontur,W.S., Tsodikov,O.V., Shkel,I. and Record,M.T. Jr (2013) Probing the protein-folding mechanism using denaturant and temperature effects on rate constants. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 16784–16789.
30. Ruff,E.F., Kontur,W.S. and Record,M.T. Jr (2015) Using solutes and kinetics to probe large conformational changes in the steps of transcription initiation. *Methods Mol. Biol.*, **1276**, 241–261.
31. Kontur,W.S., Saecker,R.M., Davis,C.A., Capp,M.W. and Record,M.T. Jr (2006) Solute probes of conformational changes in open complex ( $RP_o$ ) formation by *Escherichia coli* RNA polymerase at the  $\lambda P_R$  promoter: evidence for unmasking of the active site in the isomerization step and for large-scale coupled folding in the subsequent conversion to  $RP_o$ . *Biochemistry*, **45**, 2161–2177.
32. Kontur,W.S., Capp,M.W., Gries,T.J., Saecker,R.M. and Record,M.T. Jr (2010) Probing DNA binding, DNA opening, and assembly of a downstream clamp/jaw in *Escherichia coli* RNA polymerase- $\lambda P_R$  promoter complexes using salt and the physiological anion glutamate. *Biochemistry*, **49**, 4361–4373.
33. Guinn,E.J., Pegram,L.M., Capp,M.W., Pollock,M.N. and Record,M.T. Jr (2011) Quantifying why urea is a protein denaturant, whereas glycine betaine is a protein stabilizer. *Proc. Proc. Natl. Acad. Sci. U.S.A.*, **108**, 16932–16937.
34. Guinn,E.J., Schwinefus,J.J., Cha,H.K., McDewitt,J.L., Merker,W.E., Ritzer,R., Muth,G.W., Engelsjerd,S.W., Mangold,K.E., Thompson,P.J. et al. (2013) Quantifying functional group interactions that determine urea effects on nucleic acid helix formation. *J. Am. Chem. Soc.*, **135**, 5828–5838.
35. Diehl,R.C., Guinn,E.J., Capp,M.W., Tsodikov,O.V. and Record,M.T. Jr (2013) Quantifying additive interactions of the osmolyte proline with individual functional groups of proteins: comparisons with urea and glycine betaine, interpretation of m-values. *Biochemistry*, **52**, 5997–6010.
36. Knowles,D.B., Shkel,I.A., Phan,N.M., Sternke,M., Lingeman,E., Cheng,X., Cheng,L., O'Connor,K. and Record,M.T. Jr (2015) Chemical interactions of polyethylene glycols (PEGs) and glycerol with protein functional groups: applications to effects of PEG and glycerol on protein processes. *Biochemistry*, **54**, 3528–3542.
37. Balauff,A., Villa,E., Mahadevan,L. and Schulten,K. (2004) Structural basis for cooperative DNA binding by CAP and lac repressor. *Structure*, **12**, 123–132.
38. Wang,A.C., Revzin,A., Butler,A.P. and von Hippel,P.H (1977) Binding of *E. coli* lac repressor to non-operator. *Nucleic Acids Res.*, **4**, 1579–1593.
39. Hong,J., Capp,M.W., Saecker,R.M. and Record,M.T. Jr (2005) Use of urea and glycine betaine to quantify coupled folding and probe the burial of DNA phosphates in lac repressor-lac operator binding. *Biochemistry*, **44**, 16896–16911.
40. Capp,M.W., Pegram,L.M., Saecker,R.M., Kratz,M., Riccardi,D., Wendorff,T., Cannon,J.G. and Record,M.T. Jr (2009) Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values. *Biochemistry*, **48**, 10372–10379.
41. Chuprina,V.P., Rullmann,J.A., Lamerichs,R.M., van Boom,J.H., Boelens,R. and Kaptein,R. (1993) Structure of the complex of lac repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. Mol. Biol.*, **234**, 446–462.
42. Kalodimos,C.G., Bonvin,A.M., Salinas,R.K., Wechselberger,R., Boelens,R. and Kaptein,R. (2002) Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator O1 through alternative conformations of its DNA-binding domain. *EMBO J.*, **21**, 2866–2876.
43. Koh,J., Shkel,I., Saecker,R.M. and Record,M.T. Jr (2011) Nonspecific DNA binding and bending by  $HU\alpha\beta$ : interfaces of the three binding modes characterized by salt-dependent thermodynamics. *J. Mol. Biol.*, **410**, 241–267.
44. Tsodikov,O.V., Saecker,R.M., Melcher,S.E., Levandoski,M.M., Frank,D.E., Capp,M.W. and Record,M.T. Jr (1999) Wrapping of flanking non-operator DNA in lac repressor-operator complexes: implications for DNA looping. *J. Mol. Biol.*, **294**, 639–655.
45. Sengupta,R., Pantel,A., Cheng,X., Shkel,I., Peran,I., Stenzoski,N., Raleigh,D.P. and Record,M.T. Jr (2016) Positioning the intracellular salt potassium glutamate in the Hofmeister series by chemical unfolding studies of NTL9. *Biochemistry*, **55**, 2251–2259.
46. Lambert,D. and Draper,D.E. (2012) Denaturation of RNA secondary and tertiary structure by urea: simple unfolded state models and free energy parameters account for measured m-values. *Biochemistry*, **51**, 9014–9026.
47. Lohman,T.M., DeHaseth,P.L. and Record,M.T. Jr (1978) Analysis of ion concentration effects of the kinetics of protein-nucleic acid interactions. Applications to the lac repressor-operator interactions. *Biophys. Chem.*, **8**, 281–294.
48. Spronk,C.A., Bonvin,A.M., Radha,P.K., Melacini,G., Kaptein,R. and Boelens,R. (1999) The solution structure of lac repressor headpiece 62 complexed to a symmetrical lac operator. *Structure*, **7**, 1483–1492.
49. Jayaram,B. and Jain,T. (2004) The role of water in protein-DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 343–361.
50. Cooper,A. (2005) Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics. *Biophys. Chem.*, **115**, 89–97.