Structural Characterization and Interactions of the CFTR Regulatory Region

by

Jennifer May Reta Baker

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Biochemistry
University of Toronto

© Copyright by Jennifer M.R. Baker (2009)
Abstract

The intrinsically disordered nonphosphorylated and phosphorylated R region of CFTR and its interactions with NBD1 and SLC26A3 STAS have been characterized at residue-specific resolution, primarily using NMR. Limited chemical shift dispersion indicates that the R region is intrinsically disordered in solution and that no global folding event occurs upon phosphorylation. Chemical shifts of backbone nuclei and sidechain carbons were assigned. SSP values indicate that phosphorylation acts as a structural switch, with a reduction in helical propensity in multiple nonphosphorylated R region segments.

Free nonphosphorylated and phosphorylated R region were characterized using a variety of structural probes. Fast timescale motion indicates the presence of structural contacts in many R region segments. Hydrodynamic radii are intermediate to those expected for fully folded or denatured proteins, with the phosphorylated R region being slightly more compact. The nonphosphorylated R region was further characterized, including measurements of molecular dimensions, N-H bond vector orientation and inter-residue distances from 6 spin label sites. Using these parameters as input to the program ENSEMBLE enabled calculation of a representative pool of nonphosphorylated
R region conformations, indicating the presence of transient contacts that could not be directly discerned from the input data.

Examining labeled R region with the addition of unlabeled NBD1 provided evidence that multiple segments of nonphosphorylated R region bind and are released from NBD1 with varying affinities in a highly dynamic equilibrium. Phosphorylation relieves these interactions, with the exception of limited R region interactions near S768 when NBD1 is ATP-bound. Largely similar nonphosphorylated R region residues bind both ATP-bound ΔF508 and wild-type NBD1. Addition of unlabeled R region to labeled ATP-bound NBD1 caused spectral changes indicative of a direct interaction with more than one surface or conformational changes within NBD1 that are transmitted from one binding surface to other surface(s). Binding of unlabeled SLC26A3 STAS domain to labeled phosphorylated R region was also monitored and indicated that similar R region segments bind NBD1 and STAS, suggesting a direct competition between these two domains for binding. A model is proposed where the R region acts as a regulatory hub, integrating interactions with a variety of partners to regulate channel function.
Acknowledgements

There are so many people who have helped me with this thesis work and throughout graduate school – it is hard to know where to begin! Early on Rhea Hudson, Voula Kanelis, Patrick Finerty, Andy Jack, Hong Lin, and Andrew Chong taught me how to work in the wet lab after I arrived barely having used a pipette! Patrick helped me with LINUX and numerous computer-related questions and I was Voula's NMR data processing and analysis protégé. Irina Bezsonova, Negah Fatemi, Joe Marsh, and I learned the ropes together in a lot of our classes. The expertise of Ranjith Muhandarim and Lewis Kay at the NMR centre has been invaluable, and Ranjith helped me countless times in setting up experiments, data processing, and with NMR in general. I’ve had many excellent conversations about disordered proteins with Tanja Mittag, Micke Borg, Joe, and Elliott Stollar. The CFTR clan, starting out as ‘Charlie’s Angels’ with Rhea and Voula, and more recently also with Andrew and Jennifer Dawson has been a great working group and also a lot of fun! I’ve really enjoyed working in the lab and have met such great friends.

Many people have directly helped me with data presented in this work. Rhea purified much of the NBD1 used in the interaction experiments (thank you!). James Choy and I did the first resonance assignment of nonphosphorylated R region in parallel so that we could compare notes in assigning such a large disordered protein. Flemming Hansen helped me analyze relaxation rate data using his program Fuda. Laura de la Cruz, a summer student in the lab, did some of the initial paramagnetic relaxation enhancement experiments and Hong helped with the purification of single cysteine mutants used for PRE experiments. SAXS experiments on nonphosphorylated R region were generously carried out by Alexander Grishaev in Ad Bax’s lab at the NIH. Michael Dorwart, Patrick Thibodeau, and Philip Thomas at the University of Texas Southwestern have been excellent collaborators and shared their expertise with the STAS domain and NBD1. My graduate work has been supported by scholarships from NSERC, Restracomp, and the Canadian Cystic Fibrosis Foundation, as well as operating grants to Julie from CIHR, the Canadian Cystic Fibrosis Foundation, and the US Cystic Fibrosis Foundation.
I would like to thank Julie Forman-Kay for her guidance and support. Julie has really
been a true mentor, scientific and otherwise, rather than just a supervisor. I would also
like to thank my graduate committee Daniela Rotin and Alan Davidson for guidance
throughout my degree. Finally, I would like to thank my family, including my parents, my
sister Lorraine and brother-in-law Greg, brother Ryan, grandparents, aunts and uncles,
cousins, and especially my husband Jeff for their support throughout my work.
# Table of Contents

Abstract................................................................................................................................. ii

Acknowledgements................................................................................................................. iv

Table of Contents..................................................................................................................... vi

List of Tables............................................................................................................................... xi

List of Figures.............................................................................................................................. xii

List of Abbreviations................................................................................................................... xv

Chapter 1 : Introduction ............................................................................................................. 1

1.1 Disordered proteins ................................................................................................................. 2

1.1.1 Characteristics of disordered proteins ............................................................................. 2

1.1.2 Biological role of disorder ............................................................................................... 2

1.1.3 Approaches to study disordered proteins ....................................................................... 5

1.2 NMR techniques .................................................................................................................... 7

1.2.1 NMR theory ..................................................................................................................... 7

1.2.2 chemical shift .................................................................................................................... 8

1.2.3 Triple resonance assignment experiments ...................................................................... 10

1.2.4 Chemical shift analysis ..................................................................................................... 12

1.2.5 $R_2$ relaxation experiments ............................................................................................. 14

1.2.6 Heteronuclear nuclear Overhauser enhancement ......................................................... 14

1.2.7 Paramagnetic relaxation enhancement ............................................................................ 14

1.2.8 Residual dipolar couplings ............................................................................................ 15

1.3 Computation of representative disordered ensembles ......................................................... 17

1.4 CFTR .................................................................................................................................... 19

1.4.1 Cystic fibrosis and CFTR overview ............................................................................... 19

1.4.2 Nucleotide binding domains and nucleotides in channel regulation ............................. 22
1.4.3 Properties of the R region and CFTR phosphorylation ..............25
1.4.4 R region interactions .................................................................27
1.5 Rationale ..........................................................................................30

Chapter 2: Spectral Assignment of the Nonphosphorylated and Phosphorylated R Region ..........................................................................................32
2.1 Introduction ......................................................................................33
2.2 Experimental procedures .................................................................33
  2.2.1 Protein expression and purification of the R region .................33
  2.2.2 PKA phosphorylation .................................................................34
  2.2.3 NMR experiments .................................................................35
  2.2.4 SSP calculations .................................................................35
2.3 Results ............................................................................................36
  2.3.1 Comparison of F833 and L833 R region .........................36
  2.3.2 R Region from native and denaturing purifications is identical ....36
  2.3.3 The R region is disordered independent of phosphorylation ....37
  2.3.4 Nonphosphorylated and phosphorylated R region assignment ...40
  2.3.5 Phosphorylation reduces R region helicity ..............................41
2.4 Discussion ......................................................................................42

Chapter 3: Structural Characterization of the Free R Region .................44
3.1 Introduction ......................................................................................45
3.2 Experimental procedures .................................................................45
  3.2.1 Protein preparation .................................................................45
  3.2.2 Nanosecond-picosecond timescale relaxation experiments .......46
  3.2.3 Hydrodynamic radius experiments ....................................46
  3.2.4 Small-angle x-ray scattering .............................................47
  3.2.5 Residual dipolar couplings ....................................................47
3.2.6 Paramagnetic relaxation enhancement experiments.........................47
3.2.7 ENSEMBLE calculations ...............................................................48
3.2.8 CF-causing R region mutant chemical shifts ....................................49
3.3 Results ..............................................................................................49
3.3.1 Phosphorylation reduces R region structural contacts.......................49
3.3.2 Molecular size distribution ..............................................................51
3.3.3 R region bond angle orientation .....................................................52
3.3.4 Transient tertiary contacts throughout the nonphosphorylated R region........................................................................53
3.3.5 Hydrodynamic and 2° structure properties of calculated ensembles..................................................................................57
3.3.6 R region conformational ensemble ..................................................60
3.3.7 Characterizing R region CF-causing mutations ...............................64
3.4 Discussion ..........................................................................................67

Chapter 4: The R Region Interacts with NBD1 Predominantly Via Multiple Transient Helices ........................................................................73
4.1 Introduction .........................................................................................74
4.2 Experimental procedures ......................................................................74
  4.2.1 Protein preparation .........................................................................74
  4.2.2 $^{15}$N/$^{13}$C labeled R region interaction experiments with unlabeled wild-type and ΔF508 mNBD1 ...............................75
  4.2.3 $^{15}$N labeled wild-type mNBD1 interaction experiments with unlabeled R region ..........................................................76
4.3 Results ..................................................................................................76
  4.3.1 NBD1 predominantly binds nonphosphorylated R region ...............76
  4.3.2 Nonphosphorylated R region binds ΔF508 NBD1 similarly to wild-type NBD1 ........................................................................76
  4.3.3 NBD1 spectral changes upon nonphosphorylated R region binding .................................................................................81
4.4 Discussion ........................................................................................................84

Chapter 5: Characterizing the SLC26A3 STAS Domain and Interactions with the R region.........................................................................................................................91

5.1 Introduction .......................................................................................................92

5.2 Experimental procedures ................................................................................93

5.2.1 Protein expression and purification ..............................................................93

5.2.2 Comparison of $^{15}$N-labeled Wild-type and $\Delta$Y526/7 STAS domains ..........................................................94

5.2.3 $^{15}$N/$^{13}$C labeled R region interaction experiments with unlabeled wild-type STAS domain ........................................................................94

5.3 Results ...............................................................................................................94

5.3.1 $\Delta$Y526/7 mutation does not change STAS domain structure..............94

5.3.2 Phosphorylated R region interaction with STAS domain .................96

5.4 Discussion ........................................................................................................97

Chapter 6: Summary and Future Directions ..........................................................100

6.1 Summary .........................................................................................................101

6.2 Future Directions ............................................................................................103

6.2.1 Further structural characterization of the phosphorylated R region ..........................................................103

6.2.2 Mapping the binding interface of human R region and human NBD1 ........................................................................104

6.2.3 Comparison of wild-type and $\Delta$F508 NBD1 binding to R region 105

6.2.4 Characterizing the dynamic R region – NBD1 complex .....................106

6.2.5 R region binding to NBD2 and NBD1/NBD2 dimerization ............106

6.2.6 R region binding to N$_i$ and C$_i$ elbow helices ..................................107

6.2.7 Interactions of R region (817-838) negatively charged residues 109

6.2.8 R region interactions with SLC26A3 STAS domain .......................110

6.3 Conclusion ......................................................................................................111
List of Tables

Table 1.1: Physical properties of select stable biological isotopes. ................................8
List of Figures

**Figure 1.1:** Schematic of the disordered protein conformational ensemble. ..................2

**Figure 1.2:** Schematic of SAXS data collection for a protein sample .........................7

**Figure 1.3:** Example HSQC spectra of folded and disordered proteins .....................10

**Figure 1.4:** Sample triple resonance experiment strips for protein backbone assignment ....................................................................................................................................................12

**Figure 1.5:** Measurement of RDCs using solution NMR ........................................16

**Figure 1.6:** Schematic of the program ENSEMBLE ............................................19

**Figure 1.7:** Schematic of CFTR .............................................................................20

**Figure 1.8:** Structural changes in the gating cycle of the ABC transporter MsbA ........22

**Figure 1.9:** NBD1/NBD2 dimer showing two composite ATP binding sites ..........23

**Figure 1.10:** NBD1-RE crystal structures showing accessible RI and RE conformations ..........................................................................................................................................................24

**Figure 1.11:** Schematic of the PKA and PKC phosphorylation consensus sites ........25

**Figure 1.12:** Schematic of CFTR interactions with SLC26A3/6 ............................29

**Figure 2.1:** L833 and F833 R region have similar structural properties ....................36

**Figure 2.2:** R region samples purified from the soluble and insoluble fractions have the same conformations ..........................................................................................................................................................37

**Figure 2.3:** R region phosphorylation .....................................................................39

**Figure 2.4:** Backbone resonance assignments for nonphosphorylated and phosphorylated R region ..........................................................................................................................................................41

**Figure 2.5:** Free R region SSP values .....................................................................42
Figure 3.1  Fast timescale motion of the free R region........................................51
Figure 3.2  Residual dipolar couplings for the nonphosphorylated R region ..........53
Figure 3.3  R region single cysteine mutant chemical shift changes.....................54
Figure 3.4  Transient secondary and tertiary contacts throughout the R region .......56
Figure 3.5  Ensemble hydrodynamic properties....................................................58
Figure 3.6  Comparison of ensemble and experimental values ............................59
Figure 3.7  Clustered R region conformers .......................................................61
Figure 3.8  R region conformational ensemble contact map ..................................63
Figure 3.9  R region sequence properties ..........................................................64
Figure 3.10 Chemical shift changes with CF-causing mutations ..........................65
Figure 3.11 Altered transient contacts with the D806G CF-causing mutation ........67
Figure 3.12 Transient contacts of R region segments with α-helical propensity........71
Figure 4.1 Interaction of the R region with wild-type NBD1 ...............................77
Figure 4.2 Analysis of R region interactions with wild-type NBD1 .......................79
Figure 4.3 Comparison of nonphosphorylated R region interactions with wild-type and ΔF508 NBD1 ................................................................................81
Figure 4.4 Nonphosphorylated R region binding to NBD1 ..................................84
Figure 4.5 Possible modes of NBD1-R region interactions .................................87
Figure 4.6 Schematic illustrating phosphorylation-induced structural changes in the R region and consequent redistribution of binding equilibria with various intramolecular regulatory interaction partners. .............................................89
Figure 5.1  Comparison of wild-type and ΔY526/7 STAS domain TROSY-HSQC spectra .................................................................95

Figure 5.2  Analysis of phosphorylated R region interactions with STAS and NBD1 ....97

Figure 5.3  Schematic illustrating phosphorylation-induced structural changes in the R region and consequent redistribution of binding equilibria with various regulatory interaction partners. ........................................................................................................98

Figure 6.1  Properties of the putative amphipathic C-telopeptide helix .........................108

Figure 6.2  Charge distribution in the R region..................................................................109

Figure 6.3  Candidate ICDc positively-charged surface for interactions with negatively-charged R region residues ..................................................................................................................110
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>ATP-binding cassette</td>
</tr>
<tr>
<td>ACBP</td>
<td>acyl-coenzyme A binding protein</td>
</tr>
<tr>
<td>CF</td>
<td>cystic fibrosis</td>
</tr>
<tr>
<td>CFTR</td>
<td>cystic fibrosis transmembrane conductance regulator</td>
</tr>
<tr>
<td>CLD</td>
<td>congenital chloride-losing diarrhea</td>
</tr>
<tr>
<td>CNS</td>
<td>crystallography and NMR system</td>
</tr>
<tr>
<td>CPD</td>
<td>Cdc4 phosphodegron</td>
</tr>
<tr>
<td>CSI</td>
<td>chemical shift index</td>
</tr>
<tr>
<td>DISPHOS</td>
<td>disorder-enhanced phosphorylation sites predictor</td>
</tr>
<tr>
<td>DSS</td>
<td>2,2-Dimethyl-2-silapentane-5-sulfonic acid</td>
</tr>
<tr>
<td>$D_t$</td>
<td>translational diffusion coefficient</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>ICD</td>
<td>intracellular domain</td>
</tr>
<tr>
<td>MCRES</td>
<td>Monte Carlo replica sampling</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamics</td>
</tr>
<tr>
<td>$\alpha$-MoRE</td>
<td>$\alpha$-helical molecular recognition element</td>
</tr>
<tr>
<td>MSD</td>
<td>membrane spanning domain</td>
</tr>
<tr>
<td>MTSN</td>
<td>S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl</td>
</tr>
<tr>
<td>NBD</td>
<td>nucleotide binding domain</td>
</tr>
<tr>
<td>NOE</td>
<td>nuclear Overhauser effect</td>
</tr>
<tr>
<td>N-tail</td>
<td>N-terminus</td>
</tr>
<tr>
<td>PKA</td>
<td>protein kinase A</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>PKC</td>
<td>protein kinase C</td>
</tr>
<tr>
<td>PFG</td>
<td>pulse field gradient</td>
</tr>
<tr>
<td>pKID</td>
<td>phosphorylated kinase inducible activation domain</td>
</tr>
<tr>
<td>Po</td>
<td>channel open probability</td>
</tr>
<tr>
<td>PRE</td>
<td>paramagnetic relaxation enhancement</td>
</tr>
<tr>
<td>pSer</td>
<td>phosphoserine</td>
</tr>
<tr>
<td>R</td>
<td>regulatory</td>
</tr>
<tr>
<td>RDC</td>
<td>residual dipolar coupling</td>
</tr>
<tr>
<td>RE</td>
<td>regulatory extension</td>
</tr>
<tr>
<td>Rg</td>
<td>radius of gyration</td>
</tr>
<tr>
<td>Rh</td>
<td>hydrodynamic radius</td>
</tr>
<tr>
<td>RI</td>
<td>regulatory insert</td>
</tr>
<tr>
<td>SAXS</td>
<td>small angle x-ray scattering</td>
</tr>
<tr>
<td>SSP</td>
<td>secondary structural propensity</td>
</tr>
<tr>
<td>STAS</td>
<td>anti-sigma antagonist domain</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction
1.1 Disordered proteins

1.1.1 Characteristics of disordered proteins

The traditional view of the protein structure-function relationship that requires proteins to fold into a fixed three-dimensional structure as a prerequisite for functionality has been challenged in recent years. A new class of intrinsically disordered proteins or segments of proteins exist as an ensemble of rapidly inter-converting heterogeneous conformations, rather than a relatively stable, homogenous conformation of a folded protein (Figure 1.1) (Wright and Dyson 1999). Individual conformations vary in their compactness and tertiary contacts and the population-weighted average conformation has only fractional secondary structure. These proteins have lower sequence complexity than folded proteins and a greater proportion of Arg, Lys, Glu, Pro, and Ser residues, with fewer Cys, Tyr, Trp, Ile, and Val, resulting in a higher hydrophilicity and net charge. (Romero, Obradovic et al. 2001).

![Figure 1.1: Schematic of the disordered protein conformational ensemble.](image)

1.1.2 Biological role of disorder

Disordered protein segments are prevalent and are involved in many regulatory processes. Prediction of disordered protein segments in sequenced genomes indicates that between 2-40% of proteins contain region(s) of at least 50 contiguous disordered
residues. These segments may be involved in higher-level regulatory processes, as they are found in a higher proportion of eukaryotic proteins (25-40%) than bacterial (4-25%) or archaeal (2-20%) proteomes (Dunker, Obradovic et al. 2000). Proteins predicted to contain disordered regions are frequently associated with regulatory processes such as differentiation, the cell cycle, and cell division, while metabolic processes are strongly associated with order (Xie, Vucetic et al. 2007). Characteristics of phosphorylated protein sequences are similar to disordered protein segments including amino acid composition, low sequence complexity, low hydrophobicity, and high charge. These characteristics were used to develop the phosphorylation site prediction software disorder-enhanced phosphorylation sites predictor (DISPHOS) (Iakoucheva, Radić et al. 2004). Consistent with this, phosphorylation is by far the most common post-translational modification found in predicted disordered segments (Xie, Vucetic et al. 2007), again suggesting an intimate relationship between disordered protein segments and regulation.

The plasticity of disordered proteins is also thought to facilitate their regulated binding to a variety of binding partners (Wright and Dyson 1999). Analysis of eukaryotic interactomes showed that proteins classified as hub proteins (having >10 interaction partners) are enriched in disordered regions compared to end proteins (having only one interaction partner) (Haynes, Oldfield et al. 2006). Intrinsically disordered proteins are proposed to have energetic advantages in molecular recognition to fine-tune the free energy of binding. Compared to a folded protein, a disordered protein in the free state possesses a higher degree of conformational freedom, such that binding causes a greater restriction in motion and a corresponding greater decrease in entropy. The enthalpic component of binding for disordered proteins may also be more favourable than folded proteins in that a larger binding interface in proportion to primary sequence length is possible. Compensation between these altered entropy and enthalpy terms could allow for increased control of specificity and affinity for disordered proteins over folded proteins that is desirable in many key cellular processes. Stabilization of interacting structural elements upon binding results in entropic penalties that reduce the affinity of the interaction, providing reversibility that is important for proteins involved in
inducible events, such as phosphorylation, and for those with multiple interaction partners (Wright and Dyson 1999).

The ensemble of interconverting conformers in disordered proteins may also impart kinetic advantages to interactions. In the proposed ‘fly-casting mechanism’ model, the presence of extended conformers gives a disordered protein an increased capture radius to form initial contacts with its binding partner. Once an initial interaction occurs, the binding partner can then be ‘reeled in’ to form further binding contacts (Shoemaker, Portman et al. 2000). The ensemble of structures in a disordered protein also increases the accessible surfaces available to form initial contacts. This is illustrated by the phosphorylated kinase inducible activation domain (pKID) of the transcription factor CREB, which exists as an ensemble of transient ‘encounter’ complexes with its binding partner, the KIX domain of the CREB binding protein, before undergoing a coupled folding and binding reaction (Sugase, Dyson et al. 2007).

Although there are many examples of proteins that are intrinsically disordered in the free state and then undergo folding transitions to the bound state upon binding (Dyson and Wright 2002; Dyson and Wright 2005), these proteins actually represent a specific case of disordered protein binding. The full spectrum of disordered protein behaviour includes systems where the disordered protein adopts a small number of conformations upon binding, where coupled folding and binding occurs for only a single helix or super secondary structural element with flanking residues remaining disordered, and where a segment between two folded binding domains remains disordered and free while the two domains are bound. An additional class of intrinsically disordered protein interactions, of which there are limited examples in the literature, is that in which disordered regions remain highly dynamic even in the bound state, and display only local ordering in the context of transient contacts with their binding partner (Tompa and Fuxreiter 2008).

Two such dynamic complexes have been examined in the literature, both involving proteins with multiple phosphorylation sites but with distinct consequences for regulation, indicating the diversity of possible mechanisms. The intrinsically disordered cyclin-dependent kinase inhibitor Sic1 has multiple suboptimal Cdc4 phosphodegron
(CPD) motifs that, when phosphorylated, bind to a single site on the Cdc4 component of an SCF ubiquitin ligase. Although the individual binding sites have weak Cdc4 binding (Borg, Mittag et al. 2007), phosphorylation at any six of these nine sites results in a dynamic equilibrium in which each site comes on and off Cdc4, allowing high affinity "switch-like" binding (Nash, Tang et al. 2001; Mittag, Orlicky et al. 2008). The subsequent ubiquitination and degradation of Sic1 triggers progression of the cell cycle through the G1-S transition (Tyers 1996; Verma, Annan et al. 1997). This behaviour is crucial, as replacing the multiple suboptimal CPDs with a single optimal site removes the switch-like behaviour and leads to chromosome instability (Nash, Tang et al. 2001).

In contrast, phosphorylation of multiple sites in the disordered N-terminal segment of the Ets-1 transcription factor acts as a "rheostat" to control autoinhibition of the ETS domain (Pufall, Lee et al. 2005). Phosphorylation within this disordered segment dampens motion of the adjacent folded domain, inhibiting its DNA binding in a graded manner. Loss of phosphorylation and/or truncation of the regulatory segment increases motion of the adjacent domain and permits DNA binding. The contrasting behaviour of Sic1 and Ets-1 serve to illustrate the range of roles that dynamic interactions involving disordered protein segments can play.

1.1.3 Approaches to study disordered proteins

Dynamic, fluctuating interaction segments are an area of recent interest, and experimental approaches for examining these systems are under active development. Many of the biophysical techniques used to study folded proteins have been adapted for the study of disordered proteins. X-ray crystallography can provide evidence of mobility by identifying regions that lack crystallographic density, and these experiments provided some of the first biophysical evidence for disordered protein segments (Bloomer 1978; Bode, Schwager et al. 1978). Far-UV CD (180-240 nm) can provide global evidence of the extent of helical or strand secondary structure by comparison of spectra to references for each secondary structure type (Kelly 1997), although this technique is qualitative and may be complicated by overlapping aromatic absorbances. While folded proteins adopt relatively compact structures, the conformer ensembles of disordered proteins are more extended. One probe of global diffusion properties uses NMR pulsed field gradient experiments, allowing determination of the hydrodynamic radius ($R_h$)
(Wilkins, Grimshaw et al. 1999). These experiments use coding and decoding gradients of various strengths. A measure of overall protein diffusion is determined by the loss of NMR signal as a given protein molecule diffuses during the timescale of the experiment and experiences different coding and decoding gradients, resulting in a loss of signal coherence. Another approach, small angle x-ray scattering (SAXS), compares the scattering profile for a protein to a buffer sample alone (Figure 1.2). For folded proteins, SAXS data analysis allows determination of the radius of gyration ($R_g$), a direct measure of molecular dimensions, as well as the mean pair distribution function $p(r)$. However, the assumptions required to perform these calculations are not valid for disordered proteins, so analysis instead focuses on the direct data obtained in the scattering profile (Marsh, personal communication). NMR translational diffusion and SAXS experiments are complementary methods to determine the distribution of molecular sizes in the disordered state ensemble with the $R_h$ being more sensitive to compact conformers and SAXS data being more sensitive to extended conformers (Choy, Mulder et al. 2002). A further measure of the distribution of distances within a molecule is with pulsed electron polarization resonance, allowing the distribution of distances between two cysteine-coupled spin labels to be determined (Jeschke and Polyhach 2007). While these techniques provide useful information about the disordered state ensemble, the most powerful biophysical measures can be obtained using site-specific NMR techniques.
**Figure 1.2:** Schematic of SAXS data collection for a protein sample

(a) The incident beam $k$ is scattered from either a reference buffer sample or a protein sample, shown in red. The intensity of the scattered beam $k_1$ is measured on the linear detector with its angular dependence $\theta$ and the difference in scattering between the protein and reference buffer samples is calculated. (b) The scattering intensity is plotted relative to the beam intensity at scattering angle zero for various scattering vectors.

### 1.2 NMR techniques

#### 1.2.1 NMR theory

NMR spectroscopy exploits the quantum mechanical principal of nuclear spin, whereby spins align when placed within a large external magnetic field. Stable isotopes of many biologically interesting nuclei have spin $\frac{1}{2}$, including those shown in table 1.1. Nuclei precess in an external magnetic field $B_0$ at their Larmour frequency, which is given by:

$$\omega = -\gamma B_0 \quad (1.1)$$

where the gyromagnetic ratio ($\gamma$) is specific to each nucleus type (see Table 1.1). For $^1$H or $^{13}$C spin $\frac{1}{2}$ nuclei with positive values of $\gamma$ there are two populations, a lower energy state that aligns with the magnetic field and a higher energy state that aligns against the magnetic field. The difference in energy ($\Delta E$) between these two states depends on Plank's constant $h$, as well as $\gamma$ and $B_0$, as in equation 1.2:

$$\Delta E = -\gamma h B_0 \quad (1.2)$$
There is a population difference between the two energy states, $N_{\text{upper}}$, the population in the upper energy state, and $N_{\text{lower}}$, the population in the lower energy state, related to $\Delta E$ and the Boltzmann distribution such that:

$$\frac{N_{\text{upper}}}{N_{\text{lower}}} = e^{-\frac{\hbar B}{kT}} \quad (1.3)$$

where $k$ is the Boltzmann's constant and $T$ the temperature. In common with other spectroscopies, spins in the lower energy state can be excited to the higher energy state when energy corresponding to $\Delta E = \hbar \nu$ is applied to the system. In the case of NMR, this occurs by means of radiofrequency waves. NMR signals are not detected when the net magnetization is aligned with $B_0$, but when radiofrequency pulses have been applied to bring magnetization into the transverse plane it can be detected by receiver coils. The population difference $N_{\text{upper}}/N_{\text{lower}}$ between spin states is small, making NMR a low sensitivity technique. Methods to improve sensitivity involve increasing the spin population difference as in equation 1.3 by increasing the magnetic field or temperature and using uniformly isotopically labeled samples because the natural abundance of many spin $\frac{1}{2}$ nuclei is low, as shown in Table 1.1.

**Table 1.1:** Physical properties of select stable biological isotopes.

<table>
<thead>
<tr>
<th>Isotope</th>
<th>Natural Abundance</th>
<th>Gyromagnetic Ratio($\gamma$) (rad s$^{-1}$T$^{-1}$)</th>
<th>Precession Frequency at 18.78T (800 MHz) ($\omega_0/\pi$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1$H</td>
<td>~100%</td>
<td>267.522 x 10$^6$</td>
<td>-800.000</td>
</tr>
<tr>
<td>$^{15}$N</td>
<td>0.37%</td>
<td>-127.126 x 10$^6$</td>
<td>81.118</td>
</tr>
<tr>
<td>$^{13}$C</td>
<td>1.1%</td>
<td>67.283 x 10$^6$</td>
<td>-201.204</td>
</tr>
</tbody>
</table>

Data adapted from table in (Levitt 2001)

### 1.2.2 Chemical shift

In a system where all nuclei are in identical chemical environments, all nuclei would precess at a uniform frequency, given by equation 1.1. For example, all protons would precess at 800MHz in an 18.78 T magnet; however nuclei in actual samples are in different chemical environments. The circulation of electrons within their electronic orbitals produces small currents that oppose the external magnetic field according to Lenz's law, producing a small local magnetic field. Each individual nucleus in the
sample will experience a slightly different net magnetic field modulated by small fields from neighbouring nuclei, either through bonding or through space. The influences of these local fields, or shielding (σ), modulate the effect of the external magnetic field and cause the nucleus to resonate at a slightly different frequency described by equation 1.4:

\[ \omega = -\gamma B_0 (1 - \sigma) \]  (1.4)

These frequency changes from the reference are small, on the order of parts per million (ppm), but form the basis of our ability to interpret all NMR experiments by allowing us to identify specific nuclei in the sample. Also, because chemical shifts report on the local environment of each nucleus, the resonance dispersion in NMR spectra indicates the amount of secondary or tertiary structure present. This is illustrated in the spectra shown in Figure 1.3, contrasting the dispersed resonances of a folded protein (a), with the overlapped resonances of an intrinsically disordered protein (b) where the various nuclei have similar average environments. The use of a proper reference frequency is critical to compare data from different spectrometers. The standard reference compound, 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), was chosen for biological NMR because it is stable under a wide variety of conditions and has a strong, upfield characteristic proton resonance that is set to 0 ppm. DSS is also used for indirect referencing in the \(^{15}\)N and \(^{13}\)C dimensions of heteronuclear experiments.
Figure 1.3: Example HSQC spectra of folded and disordered proteins

Heteronuclear Single Quantum Coherence (HSQC) spectra for (a) the folded Src SH2 domain and (b) the disordered CFTR regulatory (R) region, correlating the frequency of each amide proton in the sample with its directly attached nitrogen frequency. One peak is observed for each backbone amide, as well as Asn/Gln sidechains (circled), Arg sidechains (boxed), and Trp sidechains (red asterisk). Note that for the R region, the Trp sidechain resonance is aliased from its normal $^{15}$N chemical shift at 129.7 ppm.

1.2.3 Triple resonance assignment experiments

Triple resonance experiments are usually used to identify the residues corresponding to specific resonant frequencies in an NMR spectrum. For proteins smaller than about 25kDa in size, samples with uniform $^1$H, $^{15}$N, and $^{13}$C isotopic labeling are used (Kanelis, Forman-Kay et al. 2001). These experiments take advantage of J-couplings, or spin-spin couplings, communication through bonding electrons where a magnetic field is induced through the interaction of spins. These are effective between nuclei 1 to 3 bonds apart and average coupling values between various pairs of nuclei in a protein are known. Triple resonance experiments typically begin with the net magnetization of the proton spins: protons have the highest $\gamma$ of the biological isotopes (Table 1.1) and hence the largest net magnetization (equation 1.3). Experiments use a series of radiofrequency pulses to effect energy transitions for $^1$H, $^{15}$N, or $^{13}$C nuclei. Known J-coupling values between various nuclei pairs are exploited to direct net magnetization from one type of nucleus to another (e.g. from $^{13}$C$^\alpha$ to $^1$H$^N$ via the 2-bond coupling). Many of these assignment experiments are 3-dimensional, although the same principles
apply to further dimensions. In the course of the experiment, chemical shifts are recorded for combinations of $^1H$, $^{15}N$, or $^{13}C$ nuclei that are in a known bonding arrangement with respect to each other. One common assignment experiment, the CBCACONNH, correlates the resonance frequencies of $^{15}N$ and $^1H$ for residue $i$ with $^{13}C^\alpha/^{13}C^\beta$ frequencies for the preceding residue $(i-1)$ in three dimensions ($^{15}N_i$, $^1H^N_i$, $^{13}C^\alpha_(i-1)/^{13}C^\beta_(i-1)$). Comparing this experiment with its partner experiment, the HNCACB (correlates the resonant frequencies of both $^{15}N_i$, $^1H^N_i$, $^{13}C^\alpha_(i-1)/^{13}C^\beta_(i-1)$ and $^{15}N_i$, $^1H^N_i$, $^{13}C^\alpha_i/^{13}C^\beta_i$), allows us to map the resonances corresponding to consecutive residues in the primary sequence of the protein (Figure 1.4) (Sattler, Schleucher et al. 1999). An important second pair of experiments for sequential assignment of disordered protein sequences uses the greater intrinsic chemical shift dispersion for $^{13}C'$ (carbonyl) resonances over $^{13}C^\alpha$ and $^{13}C^\beta$ resonances. This pair of experiments includes the HNCO (correlating $^{15}N_i$, $^1H^N_i$, $^{13}C'_(i-1)$ only) with the HN(CA)CO (correlating both $^{15}N_i$, $^1H^N_i$, $^{13}C^\alpha_(i-1)$ and $^{15}N_i$, $^1H^N_i$, $^{13}C'_{(i)}$) (Sattler, Schleucher et al. 1999). These two pairs (four experiments total) can be analyzed using their common $^{15}N$ and $^1H^N$ shifts to simultaneously follow $^{13}C^\alpha/^{13}C^\beta$ and $^{13}C'$ correlations down the primary sequence. It is possible to identify some amino acid types, such as Ser, Thr, Gly, and Pro, through their characteristic $^{13}C^\alpha$ and $^{13}C^\beta$ values in these experiments. Additional experiments such as the CCC-TOCSY-NNH can be used to correlate all aliphatic carbon sidechain resonances with the $^{15}N_i$ and $^1H^N_i$ of the following residue to identify most other residue types (Kanelis, Forman-Kay et al. 2001).
Figure 1.4: Sample triple resonance experiment strips for protein backbone assignment

Strips are from an HNCACB experiment (residues i-1, i) and from a CBCACONNH experiment (residue i-1 only) with positive peaks (negative peaks) in black (red), showing how assignment experiments “walk” down the protein backbone. Strips display correlations with $^{13}$C frequencies of residue (i-1) or residues (i-1, i) as indicated. Black lines indicate peaks on the same plane common to both experiments, indicating they originate from residue i-1. Red lines indicate sequential assignment connecting residues from adjacent strips on different nitrogen planes.

1.2.4 Chemical shift analysis

The chemical shifts of nuclei are determined by their local chemical environment. Therefore, once resonance frequencies have been assigned to specific residues, they also provide information about secondary structure content. The chemical shift index (CSI) calculation compares these experimental values to reference chemical shift
values for nuclei in each amino acid type in a random coil conformation (Wishart and Sykes 1994). CSI is a binary scale, with an α–helix assigned where at least 4 sequential nuclei have chemical shifts different from random coil in the same direction (positive values for $^{13}\text{C}^\alpha$) and β–strand is assigned for 4 sequential nuclei in the opposite direction (negative values for $^{13}\text{C}^\alpha$). A simple consensus of calculations for different types of nuclei is reached by taking the secondary structure predicted by the majority of nuclei types (the best predictors are $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}'$, and $^1\text{H}^\alpha$ nuclei) (Wishart and Sykes 1994; Wishart and Nip 1998). A more informative method for intrinsically disordered proteins is to use the secondary structure propensity (SSP) calculation (Marsh, Singh et al. 2006). This method uses not only the random coil reference chemical shifts, but also references for each amino acid type in fully formed α-helical or β-strand secondary structure. The calculation uses a sliding window and compares experimental chemical shifts to these reference values. The fractional population of α- or β- structure is determined by comparing the relative experimental chemical shift difference from random coil versus fully formed α-helix or β-strand values. Importantly, calculations for each residue within the window are weighted for each residue type by the magnitude of the chemical shift change expected when secondary structure is formed. The calculation also combines values for various nuclei into a common scale and uses a weighted average so that nuclei such as $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, and $^1\text{H}^\alpha$ with larger deviations from random coil are given higher weighting factors than other nuclei such as $^1\text{H}^N$ or $^{15}\text{N}$. Continuous positive values are indicative of α-helical structure, while continuous negative values are indicative of extended structure. SSP values range from +1.0, representing stable α-helical structure, to -1.0, representing stable β-structure, with 0 representing random coil. An SSP value of +0.3, for example, represents a 30% population of α-helical conformers at a particular residue, while a value of -0.1 indicates a 10% population of β-structure at another residue. For disordered proteins, SSP is particularly useful because it is a graded scale, unlike CSI which identifies fully formed secondary structure in a binary fashion, and is able to identify fractional populations of conformers with secondary structure within the primary sequence.
1.2.5 $R_2$ relaxation experiments

In a magnetic field, excited nuclear spin states decay back to the ground state through relaxation processes, the most important being dipole-dipole interactions with other nuclei. These interactions are generally averaged in solution state NMR where no dipole-dipole couplings are observed, but still play an important role in a time dependent sense by causing relaxation. $R_2$ relaxation experiments give a measure of the mobility of various nuclei on a fast timescale and have most commonly been applied to the N-H bond vectors in a protein to determine the flexibility of the protein backbone. A completely random coil protein with no tertiary contacts would be expected to have relatively uniform relaxation rates throughout its primary sequence, with reduced rates near the N- and C- termini (Klein-Seetharaman, Oikawa et al. 2002). Instead, disordered proteins have clusters of residues that exhibit increased $R_2$ relaxation rates, indicative of restricted motion by preferential transient contacts within these clusters (Klein-Seetharaman, Oikawa et al. 2002). This makes $R_2$ relaxation rates a valuable source of information about the formation of secondary and tertiary contacts within the disordered state ensemble.

1.2.6 Heteronuclear nuclear Overhauser enhancement

Another commonly measured parameter with utility for understanding fast timescale motions of proteins is the heteronuclear nuclear Overhauser enhancement (NOE), commonly applied to $^1$H-$^{15}$N bond vectors. This parameter is very sensitive to the correlation time of the molecule, with values ranging from -3.6 to 0.82 on a 500 MHz spectrometer (Kay, Torchia et al. 1989). On this scale, negative NOE values occur for the most mobile residues, while high positive values are associated with lower mobility (Ahmed, Bamm et al. 2007; Gall, Xu et al. 2007).

1.2.7 Paramagnetic relaxation enhancement

An application of $R_2$ relaxation rates that is invaluable in the study of disordered proteins is the use of paramagnetic relaxation enhancement (PRE) to derive through-space distance restraints for residue pairs throughout the protein. In these experiments, a single paramagnetic probe is coupled to the protein either through an ATCUN Cu$^{2+}$-Ni$^{2+}$ binding motif at the protein N-terminus (Donaldson, Skrynnikov et al. 2001; Marsh,
Neale et al. 2007) or by engineering single-site cysteine mutants able to covalently bind a nitrooxide spin label such as S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methylmethanesulfonothioate (MTSL) (Gillespie and Shortle 1997). Both Cu$^{2+}$ and MTSL are paramagnetic agents that increase the relaxation rates for nuclei in residues within approximately 20-25 Å of the spin label. These effects are sensitive to the distance $r$ between the nucleus and the spin label and have a $1/r^6$ dependence. A ratio of peak intensities between a spectrum where the spin label is oxidized and active to one where it is reduced and inactive can then be used to calculate distances between the spin label and various nuclei. With spin labels such as MTSL, single cysteine mutants can be made throughout the protein primary sequence, allowing distances to be calculated relative to various points in the protein primary sequence. PRE measurements provide powerful distance restraints between residues throughout the protein that can be used to characterize the disordered protein conformational ensemble.

1.2.8 Residual dipolar couplings

Although dipolar couplings are generally averaged in isotropic solution so that they are not observed (section 1.2.5), it is possible to introduce a small degree of net alignment, resulting in a small degree of anisotropy and the ability to measure a small residual dipolar coupling (RDC) (Tjandra and Bax 1997) (Figure 1.5a). These experiments are powerful because they contain angular information about specific bond vector types in a protein with respect to a common reference, the external magnetic field (Figure 1.5b). A variety of partial alignment media have been used to measure RDCs including Pf1 phage (Clore, Starich et al. 1998; Hansen, Mueller et al. 1998), strained polyacrylamide gels (Sass, Musco et al. 2000; Tycko, Blanco et al. 2000), lipid bicelles (Tjandra and Bax 1997), and liquid crystalline ethylene glycol/alcohol phases (Ruckert and Otting 2000), with the choice of alignment media depending on the experimental conditions. In practice, RDCs are typically measured by performing a pair of experiments in the absence and presence of alignment media to measure scalar J-couplings alone and then to measure the sum of the dipolar and scalar J couplings. Taking a difference between these two values gives the dipolar coupling. In folded proteins, RDC values are interpreted by considering the entire molecule in a single alignment with respect to the magnetic field. Information about the orientation of specific bond vectors in the
protein is then derived with respect to that common alignment. In disordered proteins, this interpretation is complicated by the significant dynamics present in the ensemble.

**Figure 1.5: Measurement of RDCs using solution NMR**

(a) Media such as Pf1 phage (grey rods) introduce a small net orientational alignment of the protein (shown in red), in this case by steric hindrance of protein tumbling. (b) Bonds between the nuclei probed, N-H bonds for example, have RDC values that have orientational dependence with respect to a common reference, the $B_0$ field shown with a blue arrow.

The first unfolded protein system examined using RDCs, a mutant of staphylococcal nuclease in 8M urea, had non-zero RDC values that were unexpected for a disordered protein. Similar RDC values between native and denatured protein were taken as an indication of native-like topology in the unfolded state (Shortle and Ackerman 2001). Subsequently, simulated RDCs for random flight chains indicated that non-zero values are expected for unfolded proteins (Louhivuori, Paakkonen et al. 2003). Later, RDC values were found to be correlated with fractional secondary structure propensity for unfolded proteins, with negative RDC values observed in segments with helical propensity and positive values observed in segments with strand propensity for a
sample oriented in a compressed polyacrylamide gel (Mohana-Borges, Goto et al. 2004). A simple model was proposed based on the net alignment of these secondary structural features with respect to the magnetic field and the N-H bond vector orientation within these segments. Changes in the sign of RDC values are then indicative of a change in preferential secondary structure type (Mohana-Borges, Goto et al. 2004). RDCs have since been analyzed in more detail, where experimental values were compared with synthetic RDCs calculated based on the protein sequence. These methods use a pool of conformers from a simple statistical coil model and globally align each conformer with respect to the magnetic field (Bernado, Blanchard et al. 2005; Jha, Colubri et al. 2005). Results show reasonable agreement with experimental values; however they are quite computationally intensive and require ensembles of 100,000 and 5,000 conformations, respectively. A more recent method uses local alignment of overlapping 15-residue segments within individual conformers to calculate synthetic RDCs with a much smaller conformer pool (Marsh, Baker et al. 2008). Differences in experimental RDC values to those calculated based on secondary structure can provide evidence for long range contacts (Bernado, Bertocci et al. 2005).

1.3 Computation of representative disordered ensembles

Having probed a number of experimental parameters for a disordered protein, one would like to incorporate these data into an explicit representation of the disordered protein conformational ensemble. To accomplish this, experimental parameters must be collectively satisfied by a pool of conformers since they cannot be satisfied simultaneously in a single conformation; for example distance restraints would lead to overly compact structures compared to hydrodynamic data (Gillespie and Shortle 1997). There are two main methods to obtain these disordered ensembles: using restrained molecular dynamics (MD) simulations or a Monte Carlo sampling method to optimize a conformer pool (Mittag and Forman-Kay 2007; Vendruscolo 2007), or with calculations using the program ENSEMBLE (Choy and Forman-Kay 2001; Marsh, Neale et al. 2007; Marsh and Forman-Kay 2008).

Distances derived from PRE experiments have been used with the Monte Carlo replica sampling (MCRS) method and with restrained MD simulations to enforce agreement
between experimental data and the ensemble average of a number of non-interacting replicas of the molecule, typically 20, that are simulated in parallel (Lindorff-Larsen, Kristjansdottir et al. 2004; Dedmon, Lindorff-Larsen et al. 2005). In MCRES, each Monte Carlo simulation modifies a random residue of each replica. The agreement between the ensemble averaged restraints over the replicas is compared to experimental restraints from PRE experiments, and evaluated according to the Metropolis criterion (Lindorff-Larsen, Kristjansdottir et al. 2004). An alternate approach uses restrained MD simulations with a PRE data-derived energy term in addition to the standard MD force field in replica simulations (Dedmon, Lindorff-Larsen et al. 2005).

These techniques have been pioneered by the Vendruscolo group and used to study a variety of intrinsically disordered or denatured proteins including guanidine hydrochloride denatured bovine acyl-coenzyme A binding protein (ACBP) (Lindorff-Larsen, Kristjansdottir et al. 2004), acid-denatured ACBP (Kristjansdottir, Lindorff-Larsen et al. 2005), the unfolded state of Δ131Δ staphylococcal nuclease (Francis, Lindorff-Larsen et al. 2006), and α-synuclein (Dedmon, Lindorff-Larsen et al. 2005).

Representative conformational ensembles can also be determined by a second approach, the program ENSEMBLE, developed in the Forman-Kay group (Choy and Forman-Kay 2001; Marsh, Neale et al. 2007; Marsh and Forman-Kay 2008). This program enables the use of a variety of experimental restraints including chemical shifts, $^3$J-couplings, NOEs, PREs, $R_2$ relaxation rates, SAXS scattering data, $R_h$, solvent accessible surface area, and RDCs. The program (Figure 1.6) begins with a large pool of conformers, that may be biased or unbiased by experimental data, calculated using programs such as TraDES (Feldman and Hogue 2000), Rosetta (Rohl, Strauss et al. 2004), and/or CNS (Brunger, Adams et al. 1998). Conformers are chosen from the pool to collectively satisfy the experimental restraints and optimized using a Monte Carlo energy minimization. The final ensemble of conformers from a given run is then modified using the experimental restraints and these conformers are included in a new input pool into the ENSEMBLE energy minimization in an iterative fashion, improving the fit of the representative ensemble to the experimental restraints. This program has the advantage of being able to incorporate a variety of types of experimental probes in calculations. With the large number of experimental probes of disordered proteins
possible, the ENSEMBLE method that can utilize these restraints to calculate a representative ensemble is a very powerful tool.

![Figure 1.6: Schematic of the program ENSEMBLE](image)

Solid lines indicate the program flow during a given minimization. Dashed lines indicate iterative ENSEMBLE runs where a set of conformers from one minimization is modified and used as inputs for the next minimization run.

1.4 CFTR

1.4.1 Cystic fibrosis and CFTR overview

CFTR is the product of the gene that is defective in cystic fibrosis (CF). CF is the most common inherited disease in Caucasian populations, with an incidence of approximately 1 in 2500 live births (Ratjen and Doring 2003). There are currently 1556 known mutations of CFTR (CFTR mutation database, February 5, 2008) with the most common, ΔF508, found in ~70% of North American CF alleles (Kerem, Rommens et al. 1989). The ΔF508 mutation is primarily a folding defect, causing most of the protein to be retained in the endoplasmic reticulum rather than trafficked to the cell surface (Welsh, Denning et al. 1993). The ΔF508 CFTR protein that does reach the cell surface shows reduced activity compared to wild-type CFTR (Wang, Zeltwanger et al. 2000). Expressed in epithelial cells, CFTR forms a Cl⁻ channel allowing ions to move down the electrochemical gradient. The loss of functional CFTR channels removes this Cl⁻ current and increases Na⁺ absorption, dehydrating the cell surface and leading to viscous secretions in the airways, pancreatic ducts, and intestine. Clinical
manifestations include chronic lung disease with airway obstruction and colonization by pathogens such as *Pseudomonas aeruginosa*, male infertility, and pancreatic insufficiency in 85-90% of CF patients (Ratjen and Doring 2003).

**Figure 1.7:** Schematic of CFTR

Schematic model of the ABC transporter CFTR in the PKA phosphorylated, ATP-bound state with the 12 membrane spanning domains numbered and the two membrane spanning domains (MSD1 and MSD2) indicated. The N- and C-terminal intracellular domains (Nt ICD and Ct ICD) are shown in dark blue. The two nucleotide binding domains (NBD1 and NBD2) are shown in light blue and aqua, respectively. The regulatory insert (RI) and regulatory (R) region are both shown as red coils rather than defined globular domains to reflect their disordered nature. PKA phosphorylation sites are also indicated in red. The putative N- and C-terminal elbow helices (Nt elbow helix and Ct elbow helix), similar to those found in the Sav1866 (Dawson and Locher 2006) and MsbA (Ward, Reyes et al. 2007) crystal structures, are shown in orange/yellow.
CFTR is a member of the ATP-binding cassette (ABC) transporter superfamily of proteins, found in species from archaea to humans. Like other ABC transporters, CFTR contains two membrane spanning domains (MSDs) and two nucleotide binding domains (NBDs) (Figure 1.7). X-ray structures solved for two full length ABC transporters with homology to CFTR, Sav1866 (Dawson and Locher 2006) and MsbA (Ward, Reyes et al. 2007) (Figure 1.8), show the existence of continuous helices extending from the transmembrane segments into the cytoplasm and connected by short helices to form the intracellular domains (ICDs). Unique among ABC transporters, CFTR is a channel rather than a transporter and has two intrinsically disordered inserted sequences, the ~30 residue regulatory insert (RI) within the NBD1 primary sequence and the much larger ~190 residue regulatory (R) region (Figure 1.7). Under normal conditions, CFTR channel activity is regulated by nucleotide binding and hydrolysis at the NBDs, and by protein kinase A (PKA) phosphorylation at multiple sites, primarily within the R region, with one site in the RI. These cytoplasmic regulatory events are relayed to the MSDs, likely via the ICDs, allowing the pore to conduct chloride ions. MsbA structures solved in multiple states (Figure 1.8) including nucleotide bound, apo ‘open’ and apo ‘closed’ (Ward, Reyes et al. 2007) highlight the large structural changes present within the gating cycle of MsbA. These changes may be similar in CFTR, although chloride is much smaller than the lipid A and lipopolysaccharide transported by MsbA.
Figure 1.8: Structural changes in the gating cycle of the ABC transporter MsbA

Structures of full-length MsbA in three different conformations, nucleotide-bound (accession code 3B60), apo-open (3B5W), and apo-closed (3B5X) are shown, with one chain of the homodimer in red, the second in blue. The elbow helices, indicated in green in all three structures, are interfacial between the cytoplasm and plasma membrane. Figure adapted from (Ward, Reyes et al. 2007).

1.4.2 Nucleotide binding domains and nucleotides in channel regulation

NBD dimerization, facilitating ATP hydrolysis, is expected to occur in the CFTR gating cycle (Moody, Millen et al. 2002) similar to that observed in crystal structures of bacterial ABC transporters including full length MsbA and Sav1866 (Dawson and Locher 2007; Ward, Reyes et al. 2007) and the NBD of MJ0796 (Moody, Millen et al. 2002; Smith, Karpowich et al. 2002). NBD1-NBD2 dimerization occurs in a head-to-tail fashion, creating two nucleotide binding sites at the dimerization interface. Each site is formed by the Walker A, Walker B, and His loop from one nucleotide binding domain along with the ABC signature sequence from the opposing nucleotide binding domain (Figure 1.9). In CFTR, the two nucleotide binding sites are non-equivalent, with many substitutions from the canonical ABC transporter sequence for residues comprising the first binding site (Figure1.9), resulting in a site that stably binds ATP, but has very limited hydrolysis (Aleksandrov, Aleksandrov et al. 2002). The second nucleotide binding site has canonical ABC transporter residues and is a site of ATP binding and rapid hydrolysis (Aleksandrov, Aleksandrov et al. 2002). Although NBD1 and NBD2 are
each able to hydrolyze ATP at a low level in isolation, both NBDs are required for optimal ATPase activity (Kidd, Ramjeesingh et al. 2004). Channel opening has been linked to ATP binding at both nucleotide binding sites and NBD dimerization (Vergani, Lockless et al. 2005), although binding to the NBD1 binding site alone allows some activity (Berger, Ikuma et al. 2005). Channel closing has been shown to be linked to ATP hydrolysis at the second ATP binding site (Berger, Ikuma et al. 2005).

**Figure 1.9:** NBD1/NBD2 dimer showing two composite ATP binding sites

The two composite ATP-binding sites are shown with the ATP hydrolysis-competent site (red star) and the hydrolysis-incompetent site (white circle). The component residues are indicated: the signature sequence, green oval; Walker A sequence, purple circle; Walker B sequence, blue circle; and the H loop, yellow circle.

Several CFTR NBD1 crystal structures have been solved (Lewis, Buchanan et al. 2004; Lewis, Zhao et al. 2005; Thibodeau, Brautigam et al. 2005; Atwell, Conners et al. 2007), two of which are shown in Figure 1.10. Three of these structures have been of fragments containing the residues 389-673 (Lewis, Buchanan et al. 2004; Lewis, Zhao et al. 2005; Thibodeau, Brautigam et al. 2005), with a novel insert in the core ABC transporter sequence of residues 404-435, termed the regulatory insert (RI) and the regulatory extension (RE), residues ~650-673. The RE comprises the first ~20 residues of the R region. The RI and RE have increased B factors in the crystal structures and
are highly mobile, as shown by comparing the various crystal structures (Figure 1.10a, 1.10b). The position of the RE in several crystal structures interferes with NBD1/NBD2 dimerization based on the dimer seen in Sav1866 (Dawson and Locher 2006), indicating that interactions between NBD1 and the RE must be released prior to dimerization. This is supported by recent work on NBD1(389-646)/Δ(404-432), which contains the entire NBD fold with both the RI and RE removed. This protein region crystallized in both monomeric and homodimer forms (Atwell, Conners et al. 2007). Note that this segment ending at residue 646 contains the entire canonical NBD fold, with a C-terminal boundary for NBD1 at approximately residue 650 based on sequence alignments of various ABC transporters (Lewis, Buchanan et al. 2004) and functional studies defining the C-terminal NBD1 boundary around L633 (Csanady, Chan et al. 2000; Csanady, Chan et al. 2005).

**Figure 1.10:** NBD1-RE crystal structures showing accessible RI and RE conformations

(a) Structure of ATP-bound NBD1-RE (1R0X) and (b) ATP-bound F508A (containing 2 additional mutations) NBD1-RE (1XMI) showing the RI and RE (pink), the F508 sidechain (red), and ATP as ball/stick. In (a) the RE blocks the NBD1/NBD2 dimerization interface. In (b) the RI and RE are rotated and in a conformation compatible with dimerization.

The structures of several F508 mutants have been determined, including the F508S non-CF-causing variant and the F508R maturation-deficient mutant (Thibodeau, Brautigam et al. 2005) and ΔF508 and F508A mutants (Lewis, Buchanan et al. 2004; Lewis, Kearins et al. 2005). All have similar structures to wild-type NBD1-RE. However, the ΔF508 and F508A structures were all solved with second site suppressor and solubilizing mutations within the sequence that have been found to partially rescue ΔF508 cell surface expression in the context of full-length CFTR (Teem, Carson et al. 2005).
When ΔF508 NBD1 with no other mutations is studied with NMR in the solution state, the ΔF508 mutation alters the structural properties of phosphorylated NBD1-RE compared to wild-type, indicating that structural and dynamic changes occur in ΔF508 NBD1 (Kanelis, Hudson et al. 2009).

1.4.3 Properties of the R region and CFTR phosphorylation

In addition to ATP binding and hydrolysis at the NBDs, normal channel activity also requires PKA phosphorylation. The dibasic (R-R/K-X-S/T) and monobasic (R-X-S/T) PKA consensus sequences have high interspecies conservation (Riordan, Rommens et al. 1989). They are located within the disordered phosphoregulatory segments, with nine in the R region and the 10th site in the RI within the NBD1 sequence (Figure 1.11). Additionally, there are nine potential protein kinase C (PKC) recognition sites (Figure 1.11) that, when phosphorylated, increase the efficiency of phosphorylation by PKA and alone may partially activate CFTR (Chappe, Hinkson et al. 2004).

![Figure 1.11: Schematic of the PKA and PKC phosphorylation consensus sites](image)

The NBD1 core sequence (grey) and RI and R region sequences (pink) are shown with consensus PKA phosphorylation sites (red) and PKC phosphorylation sites (blue) shown below the sequence. Approximate domain boundaries are indicated.

The R region has been shown to be predominantly disordered by CD (Dulhanty and Riordan 1994; Ostedgaard, Baldursson et al. 2000) and contains 28% charged amino acids (R,K,E,D) in its primary sequence (residues 654-838). Overall, five serines (S660, S700, S737, S795, and S813) have been shown to be phosphorylated in vivo (Cheng, Rich et al. 1991), while an additional 5 sites (S422, S670, S712, S768, and S753) have been detected in vitro (Neville, Rozanas et al. 1998; Lewis, Buchanan et al. 2004). These multiple phosphorylation sites are generally additive in controlling CFTR channel.
opening, without a requirement for phosphorylation at any one specific site (Cheng, Rich et al. 1991; Chang, Tabcharani et al. 1993; Rich, Berger et al. 1993). This redundancy was further confirmed in studies deleting both phosphoregulatory segments, the RI and the RE, as these mutant channels had gating properties similar to wild-type (Csanady, Chan et al. 2005). Functional studies of full-length CFTR incorporating missense Ser→Ala mutations in PKA consensus motifs, singly or in combination, demonstrate various relative contributions of these positions to channel activity (Wilkinson, Strong et al. 1997; Vais, Zhang et al. 2004). Whereas most phosphorylation sites stimulate channel activity, Ser737 and Ser768 are inhibitory sites and substitutions at these residues result in increased channel conductance (Wilkinson, Strong et al. 1997) or Po (Vais, Zhang et al. 2004). Phosphorylation at most sites appears to increase Po by increasing the rate of channel opening (Winter and Welsh 1997; Mathews, Tabcharani et al. 1998; Vais, Zhang et al. 2004). The inhibitory phosphorylation sites may decrease Po by different means: at S737 by decreasing the rate of channel opening (Vais, Zhang et al. 2004) and at S768 by shortening the channel open time (Csanady, Chan et al. 2005).

In addition to covalent modification by phosphorylation, other R region structural changes such as proline isomerization may contribute to R region regulation of CFTR function. The cis-trans proline isomerase cyclophilin A has been shown to increase the Po of PKA-activated CFTR in the presence of ATP. Three highly conserved prolines, P740, P750, and P759, have been implicated in this response, suggesting that R region structural changes independent of PKA phosphorylation may also modulate channel activity (Xie, Zhao et al. 2000).

The stimulation of CFTR channel activity and prevention of activity in the absence of phosphorylation by the R region is complex and context-specific. The deletion of R region residues from 708-835 (Rich, Marshall et al. 1991; Ma, Zhao et al. 1997), or even only the residues 760-783 (Baldrursson, Ostedgaard et al. 2001), or 817-838 (Xie, Adams et al. 2002) produces channels that open in the absence of PKA phosphorylation, but with a lower open probability (Po) than wild-type. The phosphorylated R region has a stimulatory effect on channel activity that depends on context. The addition of exogenous R region in either the nonphosphorylated or
phosphorylated states has no effect on wild-type CFTR or CFTR containing mutated phosphorylation sites (Winter and Welsh 1997), although exogeneous nonphosphorylated R(590-858) peptide blocked wild-type CFTR in one report (Ma, Zhao et al. 1997). In another context, the addition of phosphorylated R region to a CFTR mutant (Δ708-835/S660A) that gates without phosphorylation, increases channel activity by increasing the rate of channel opening, while exogenous nonphosphorylated R region has no effect (Ma, Zhao et al. 1997; Winter and Welsh 1997). This indicates a possible context-specific role in channel stimulation for phosphorylated R region that has been shown to bind other CFTR segments (Naren, Cormet-Boyaka et al. 1999) or other proteins (Ko, Zeng et al. 2004).

Overall, global R region structural changes, particularly at the residue-specific level, are poorly understood. Any insight into R region structural changes with phosphorylation in the CFTR gating cycle would significantly inform the overall mechanism of CFTR regulation.

1.4.4 R region interactions

The R region engages in many inter- and intramolecular interactions to modulate channel activity, consistent with the observation that many other disordered protein segments also interact with multiple partners (Haynes, Oldfield et al. 2006). Within CFTR, the R region must interact with other parts of the molecule to prevent channel activation in the absence of PKA phosphorylation and be modulated with phosphorylation to support channel activity. Previous work indirectly suggests NBD1/R region interactions, either through the inhibition of PKA phosphorylation in the presence of nucleotide-bound NBD1 (Neville, Rozanas et al. 1998) or through altered ATP binding with phosphorylation (Howell, Borchardt et al. 2004). Additionally, NBD1 crystal structures solved in the presence of the RE, encompassing the first twenty residues of the R region, show a potential binding site on the NBD1 core blocking the NBD1/NBD2 dimerization interface (Lewis, Buchanan et al. 2004). There is some preliminary evidence for R region binding to NBD2 through overlay assays with either peptides or the full domain, although these studies used “R region” residues from 589-830 and “NBD2” residues from 1151-1476, making interpretation difficult in the context
of current CFTR domain boundaries (Wang, He et al. 2002). More recent work has linked R region phosphorylation to NBD1/NBD2 dimerization by monitoring the cross-linking ability of specific cysteines in a cysteine-free background (Mense, Vergani et al. 2006).

In addition to interactions with the nucleotide binding domains, the R region also modulates channel activity via interactions with two CFTR segments that may be analogous to the cytoplasmic MsbA and Sav1866 elbow helices (Figure 1.8). The CFTR N-terminus (N-tail) is approximately 75 residues long before the first transmembrane segment (Figure 1.7). These residues may be part of an elongated elbow helix that ends immediately before TM1, as in MsbA and Sav1866, or may form two helices separated by P67. The N-tail directly interacts with the R region with some phosphorylation-dependent modulation (Naren, Cormet-Boyaka et al. 1999). In the context of a CFTR hemi-channel construct (residues 1-634), interactions with the N-tail were also implicated in the unexpected increased binding of the R region under phosphorylating conditions (Chappe, Irvine et al. 2005). Interactions are primarily modulated by the acidic residues D47, E51, E54, and D58 on one face of a putative α-helix (Naren, Cormet-Boyaka et al. 1999; Cormet-Boyaka, Jablonsky et al. 2004). These interactions appear to stabilize the channel open state; the triple mutant (D47/E54/D58) shortens the channel opening duration in the context of wild-type CFTR (Naren, Cormet-Boyaka et al. 1999), as well as decreasing channel open bursts in the context of a K1250A ATP hydrolysis deficient mutant that normally has a very long channel open state (Fu, Ji et al. 2001).

There are also indications that the R region binds to the C-terminal elbow helix, a region that is analogous to similar regions in MsbA and Sav1866, and which encompasses residues between the C-terminus of the R region and the start of TM 7 (Wang, He et al. 2002; Gupta, Xie et al. 2004). Residues 838-859 in the C-terminal elbow helix are essential for binding which may also be linked to proline isomerization within the R region (Gupta, Xie et al. 2004). Although it has not been explored with the C-terminal elbow helix, a phosphorylation dependence of these interactions could potentially explain the modest R region binding noted by Chappe and coworkers to the C-terminal half of CFTR (residues 837-1480) under PKA-phosphorylating conditions (Chappe,
Irvine et al. 2005). R region interactions have thus been suggested with many of the cytoplasmic components of CFTR, including the Nt and Ct elbow helices and NBDs. Interactions with ICDn and/or ICDc (Figure 1.7) are also possible, although these have not yet been tested.

In addition to these multiple intramolecular interaction partners, the R region mediates CFTR binding to SLC26A3 and SLC26A6, two chloride/bicarbonate exchangers. Interactions with the R region occur via the conserved cytoplasmic sulphate transporter and anti-sigma antagonist (STAS) domains of SLC26A3/6 (Ko, Zeng et al. 2004) (Figure 1.12). The R region-STAS domain interaction is PKA phosphorylation-dependent (Ko, Zeng et al. 2004) or has an increased affinity with phosphorylation (Dorwart, Shcheynikov et al. 2008), and also has increased affinity when CFTR and SLC26A3/6 are co-localized via their PDZ domain-binding motifs (Ko, Zeng et al. 2004). This interaction stimulates channel activity for both CFTR and SLC26A3/6 (Ko, Zeng et al. 2004).

**Figure 1.12:** Schematic of CFTR interactions with SLC26A3/6

Phosphorylated R region (red) interacts with the STAS domain (teal) at a putative interface (grey ellipse). SLC26A3 is shown with its membrane spanning domain (MSD) and CFTR is shown with 2 membrane spanning domains (MSD1 and MSD2) (grey), 2 intracellular domains (ICD1 and ICD2) (purple), and 2 nucleotide binding domains (NBD1 and NBD2) (blue). SLC26A3 and CFTR are colocalized by their C-termini binding to a PDZ domain-containing protein (light grey ellipse).
These interactions are thought to play a role in maintaining the high bicarbonate, alkaline environment of the pancreatic duct cells that is crucial for preventing premature activation of digestive enzymes and subsequent pancreatic destruction. The stoichiometries of SLC26A3 (2Cl⁻/1HCO₃⁻) versus SLC26A6 (1Cl⁻/2HCO₃⁻) (Shcheynikov, Wang et al. 2006), if combined with differential expression of SLC26A6 in the proximal pancreatic duct versus SLC26A3 in the distal duct and chloride moving into the cell through CFTR, could explain the concentration of bicarbonate seen in normal individuals, as well as the acidic conditions present in the ducts of CF patients (Ko, Zeng et al. 2004). This model could also explain the observation that some CFTR mutations give rise to channels with normal Cl⁻ transport, but reduced HCO₃⁻ transport and correlate with pancreatic insufficiency when these mutants are present in patients; other mutations cause reduced Cl⁻ transport, but patients are still pancreatic sufficient (Choi, Muallem et al. 2001).

1.5 Rationale

Despite its key importance to CFTR regulation, the biophysical properties of the R region were largely uncharacterized when I began my thesis work. Additionally, while the R region had been shown either directly or indirectly to interact with a variety of interaction partners, both within CFTR and with other proteins, the modulation of these interactions at the molecular level was uncharacterized. The overall goal of my thesis project was to investigate the structural properties of the R region and how these mediate regulation of CFTR function. These properties were primarily investigated via NMR and are discussed in the following chapters.

Chapter 2 describes the NMR resonance assignment and secondary structural properties of nonphosphorylated and PKA-phosphorylated R region. This work forms the foundation of our ability to analyze further NMR experiments.

Chapter 3 describes further experiments to structurally characterize the nonphosphorylated and phosphorylated R region using SAXS and a variety of NMR techniques including $R_2$ relaxation rates, $R_h$ from NMR diffusion experiments, PREs, and RDCs. This information is used as restraints to calculate a representative structural
ensemble of the nonphosphorylated R region free state. Additionally, structural consequences of three CF-causing mutations in the R region are examined.

In chapter 4 I describe interactions between the R region and wild-type (WT) and ΔF508 NBD1 from the perspective of both the R region and NBD1. The phosphorylation and ATP-binding dependence of these interactions is explored and binding to WT and ΔF508 NBD1 are compared.

Chapter 5 examines structural consequences of the ΔY526/7 congenital chloride-losing diarrhea (CLD) mutation within the SLC26A3 STAS domain. Also, interactions of the phosphorylated R region with the STAS domain are probed and compared to interactions with NBD1.

Chapter 6 describes the conclusions from my thesis work. I also discuss further experiments that could be applied to further characterize the R region free state, R region interactions with NBD1 and/or NBD2, and interactions with the putative Nt and Ct elbow helices.
Chapter 2: Spectral Assignment of the Nonphosphorylated and Phosphorylated R Region

This chapter contains part of the work published in Jennifer M.R. Baker, Rhea P. Hudson, Voula Kanelis, Wing-Yiu Choy, Patrick H. Thibodeau, Philip J. Thomas, and Julie D. Forman-Kay. CFTR Regulatory (R) Region Interacts with NBD1 Predominantly Via Multiple Transient Helices. (2007) *Nature Structural and Molecular Biology*, 14, 738-745. I performed the experiments, analyzed the data, and wrote the paper. Wing-Yiu Choy contributed to data analysis of some early nonphosphorylated R region assignment experiments.
2.1 Introduction

The R region contains all but one of the PKA phosphorylation sites that, in combination with ATP binding and hydrolysis at the NBDs, regulate CFTR channel activity. In order to understand how this regulation takes place at the molecular level, characterization of both the structural properties and interactions of the R region is required. NMR is the most powerful technique for providing residue-specific information on intrinsically disordered proteins. However, in order to carry out these studies assignment of NMR resonances to specific residues in the R region is first required. Here, the assignment of both nonphosphorylated and phosphorylated R region resonances using NMR triple resonance experiments employing patterns of correlations between $^{13}\text{C}'$, as well as $^{13}\text{C}^{\alpha}$ and $^{13}\text{C}^{\beta}$, chemical shifts is described. A high proportion of the nonphosphorylated and phosphorylated R region backbone and sidechain carbon resonances were assigned, one of the longest disordered protein sequences assigned to date. With chemical shift assignments available, secondary structural propensity values (SSP values) of the R region were calculated and compared between the nonphosphorylated and phosphorylated states. The modulation of fractional secondary structural features provides a structural switch whereby R region interactions with various partners can be altered upon phosphorylation. In addition, these resonance assignments provide the basis for further studies of R region structural properties (Chapter 3), and R region interactions with NBD1 (Chapter 4) and SLC26A3 STAS domain (Chapter 5).

2.2 Experimental procedures

2.2.1 Protein expression and purification of the R region

The human CFTR (P13569) R region was expressed from a plasmid encoding a 185 residue fragment (residues 654–838) with an N-terminal, His 6x tag in the pPROEX HTb vector (Invitrogen). The Ser654 N-terminal boundary was chosen to include the PKA phosphorylation consensus residues N-terminal to the important Ser660 phosphorylation site, as well as the homologous residue to Thr654 seen to act as an N-terminal helix cap in the murine NBD1-RE crystal structure (Lewis, Buchanan et al. 2004). The C-terminal boundary was chosen to include a group of negatively charged
residues with α-helical propensity that are suggested to play a functional role in CFTR channel inhibition (Xie, Adams et al. 2002). The R region sequence carried the polymorphism Leu833 from the original CFTR cloning paper (Riordan, Rommens et al. 1989), as the WT Phe833 was deleterious for protein solubility, with a maximum solubility of 50 µM under a variety of buffer conditions tested. Comparison of HSQC spectra of Phe833 and Leu833 R region showed similar structural properties for the two proteins.

Uniformly isotopically enriched R region with either $^{15}$N labeling alone or combined $^{15}$N/$^{13}$C labeling was expressed in BL21 CodonPlus RIL cells (Stratagene) grown to OD<sub>600</sub> 0.6–0.8 and induced with 1mM isopropyl-D-thiogalactoside (IPTG) for 12-16 hours at 16°C. BL21 CodonPlus cells were used for protein expression because of the presence of the presence of 15 rare Arg codons (AGA, AGG, AGA), including 3 tandem and one triple repeat, 4 rare Ile codons (ATA), 1 rare Leu codon (CTA), and 2 rare Pro codons (CCC) in the R region DNA sequence. Additional tRNAs in the BL21 CodonPlus cells compensated for the rare Ile and Leu codons and 9 of the rare Arg codons, including the 3 tandem repeats, and greatly increased R region expression. The R region was purified from the insoluble lysate fraction using a 6 M GdmCl purification including Ni<sup>2+</sup> affinity chromatography and denaturing size exclusion chromatography with a Superdex 200 column (Pharmacia), followed by HPLC with a Jupiter 10u C4 300A reverse phase column (Phenomenex). Samples were lyophilized, rehydrated, cleaved overnight with His 6x tagged Tev protease, Ni<sup>2+</sup> affinity purified to remove the His tag and Tev protease, and exchanged into the buffer of interest. R region from the soluble fraction was also purified as above, but without HPLC purification, using non-denaturing buffers, for comparison with R region from the insoluble fraction.

2.2.2 PKA phosphorylation

Reactions were performed in 50 mM Tris (pH 7.4), 50 mM MgCl<sub>2</sub>, 50 mM ATP, 2 mM DTT at 10 µM R region protein concentration. R region was incubated for 1 hour at 37 °C after addition of 100U PKA (Promega, Madison, WI) per 300 µg of protein, then another 1 hour with an additional 100U PKA per 300 µg protein. Reactions were
stopped with the addition of 6M GdmCl and 0.1% (v/v) TFA. PKA was removed by reverse phase HPLC chromatography. Phosphorylation of specific residues and overall levels were demonstrated using electrospray mass spectrometry and NMR, to be described in further detail (sections 2.3.2, 2.3.3).

2.2.3 NMR experiments

All NMR data were collected on a Varian Inova 800MHz spectrometer at 10 °C with a room temperature, triple resonance probe with actively shielded gradients. Nonphosphorylated and highly phosphorylated R region samples at 0.25mM (in 125 mM K+ phosphate (pH 6.8), 125 mM KCl, 2 mM EDTA, 2 mM benzamidine, 2 mM DTT) were used for assignment experiments. Triple resonance assignment experiments (Sattler, Schleucher et al. 1999) were performed including CBCA(CO)NNH and HNCACB to examine Cα/Cβ chemical shifts. HNCO and HN(CA)CO experiments were invaluable in the assignment process because 13C' chemical shifts have greater dispersions than 13Cα and 13Cβ chemical shifts for disordered proteins. Additionally, CCC-TOCSY-NNH experiments were recorded to examine correlations between sidechain 13C nuclei with backbone 1H N and 15N nuclei (Kanelis, Forman-Kay et al. 2001). All spectra were referenced using the internal reference DSS (sodium 2,2-dimethyl-2-silpenta-5-sulfonate). Data were processed using NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed using NMRView (Johnson and Blevins 1994).

2.2.4 SSP calculations

The program SSP uses reference chemical shift values for each amino acid type in α-helix, β-strand, and random coil conformations and determines a percent helix or percent strand value by calculating the relative experimental chemical shift difference from random coil to that expected for stable helix or strand for each amino acid type (Marsh, Singh et al. 2006). The values are weighted for each amino acid depending on the magnitude of overall chemical shift expected for secondary structure formation. The default averaging of a five residue sliding window was used. Phosphorylated serines were excluded from the calculations because of a lack of reference chemical shift values for fully stabilized helix and strand conformations.
2.3 Results

2.3.1 Comparison of F833 and L833 R region

WT R region encompassing residues 654-838 was soluble to a maximum concentration of 50 µM, insufficient for triple resonance assignment experiments, so R region with the L833 polymorphism, soluble to ~500 µM, was used. The L833 polymorphism causes minor chemical shift changes for selected residues, however the two spectra are highly similar (Figure 2.1).

![Figure 2.1: L833 and F833 R region have similar structural properties](image)

Superimposed $^1$H-$^1$H HSQC spectra of L833 (black) and F833 (red) R region are overlaid. Spectra were recorded on an 800MHz spectrometer at 10 °C, however buffer conditions differed slightly, with L833 in 125 mM K⁺ phosphate (pH 6.8), 125 mM KCl, 2 mM DTT and F833 in 20 mM Na⁺ phosphate (pH 6.6), 200 mM NaCl, 2 mM DTT.

2.3.2 R Region from native and denaturing purifications is identical

The R region samples used in NMR experiments for structural characterization were purified under denaturing conditions from the insoluble fraction. To ensure that this
protein reflects the native state, a sample was compared to one prepared from the small proportion of R region expressed in the soluble fraction and purified under native conditions. Overlaid HSQC spectra of R region samples derived from the insoluble and soluble fractions are virtually identical (Figure 2.2), indicating that renatured R region samples were suitable for further analysis.

![Figure 2.2: R region samples purified from the soluble and insoluble fractions have the same conformations](image)

Superimposed $^1$H$^N$-$^{15}$N HSQC spectra of R region from the soluble fraction (red) and from the insoluble fraction (black) are virtually identical, indicating that they sample the same conformational ensembles.

2.3.3 The R region is disordered independent of phosphorylation

To examine changes in the structural properties of the isolated R region with phosphorylation, purified R region was phosphorylated *in vitro* with PKA catalytic subunit. Analysis by mass spectrometry indicated that phosphorylation at 8 to 9 sites of the R region was achieved. NMR $^1$H$^N$-$^{15}$N correlation spectra for nonphosphorylated R region and highly phosphorylated R region are superimposed in Figure 2.3a, showing sharp peaks, with dispersion for backbone amide proton resonances limited to values between approximately 8 and 8.7 ppm in the proton dimension. This limited dispersion
is diagnostic of disorder (vs a dispersion of ~ 7 to 10 ppm for folded proteins) and reflects the rapid inter-conversion between heterogeneous conformations in disordered proteins in which all nuclei experience similar average chemical environments (Dyson and Wright 2001). More arginine side chain resonances were observed for the phosphorylated R region, probably because interactions between phosphates and arginine guanidinium groups led to reduced solvent exchange. Phosphorylation produces marked downfield chemical shift changes for phosphoserine residues (Bienkiewicz and Lumb 1999), due to the addition of a charged phosphate (Figure 2.3). Phosphorylation does not, however, induce a global folding event as in both nonphosphorylated and phosphorylated states the observed ensemble averaged chemical shift values reflect a predominantly disordered state.
Figure 2.3: R region phosphorylation

(a) $^1$H-$^{15}$N HSQC NMR spectrum of the isolated human R region, comprising CFTR residues 654-838. Nonphosphorylated R region (black) and phosphorylated R region (red) are superimposed. Owing to the spectral width, the apparent chemical shift of the single tryptophan side chain peak is moved (aliased) to the top left corner (marked with blue asterisk). The true $^{15}$N chemical shift is 129.7 ppm. Apparent chemical shifts of the arginine side chain resonances are also aliased and appear at the lower right (dashed blue ellipse). A dashed blue rectangle marks glutamine and asparagines side chain resonances at the upper right.  

(b) Calculated $^1$H chemical shift difference between nonphosphorylated and phosphorylated R region resonances (nonphosphorylated minus phosphorylated). Red circles indicate phosphorylation sites.
2.3.4 Nonphosphorylated and phosphorylated R region assignment

Assignments were obtained for 97% of the $^1\text{H}$, $^{15}\text{N}$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, and $^{13}\text{C}'$ resonances of the nonphosphorylated R region sequence and 99% of the resonances of the phosphorylated sequence using a variety of triple resonance experiments (Figure 2.4). Residues 718–722 in the nonphosphorylated R region lack assignments because of resonance broadening (ie. loss of intensity accompanied by an increase in NMR resonance linewidth) likely due to ms-μs timescale sampling of a low population of stabilized conformations in this region. Based on analysis of resonance intensities in NMR spectra, the phosphorylated R region sample had complete phosphorylation at the previously reported PKA phosphorylation sites 660, 700, 712, 737, 753, 768, 795, and 813 (Townsend, Lipniunas et al. 1996; Neville, Rozanas et al. 1997), partial phosphorylation (approximately 60%) of Ser670 (Lewis, Buchanan et al. 2004), and ~15% phosphorylation at the novel PKA phosphorylation site Thr788. Chemical shift assignments confirmed that the largest chemical shift changes with phosphorylation occur for residues close in sequence to the phosphorylated serines (Figure 2.3b).
Figure 2.4: Backbone resonance assignments for nonphosphorylated and phosphorylated R region

Primary sequences are shown and coloured by assigned (blue) and unassigned (black) residues for (a) nonphosphorylated R region and (b) phosphorylated R region. All phosphorylated serines (red) were assigned. CFTR residue numbers indicate the residue boundaries in the construct (a). The first 5 residues (QGAME) are derived from the vector and cloning.

2.3.5 Phosphorylation reduces R region helicity

The availability of resonance assignments permits the analysis of the fractional population of secondary elements for individual residues within the R region, which was performed using the program SSP (Marsh, Singh et al. 2006) with $^{13}$Cα and $^{13}$Cβ chemical shifts as inputs (Figure 2.5). In folded proteins, consecutive residues have SSP values of +1 for α-helical structure and −1 for β-strand structure, respectively. In disordered proteins, consecutive residues with positive or negative values reflect the local fractional population in α-helical or β-strand conformations as a weighted average of the conformers present. In the nonphosphorylated R region residues 654–668, 759–764, 766–776, and 801–817 all have a greater than 5% α-helical population, with values above 20% for some residues, while 744–753 has a greater than 5% β-strand population with some values above 20%. Phosphorylation produces a global decrease in helical content, consistent with that observed in circular dichroism experiments (Dulhanty and Riordan 1994). Phosphoserine (pSer) at the N-cap or first three positions
(as for position 768) stabilizes helices, while it destabilizes within the helix or at the C-terminus (Andrew, Warwicker et al. 2002) (as for positions 700, 737, and 813). Ser660 is within a long helical segment (654–670), which shortens to 660–667 so that pSer660 becomes the N-cap residue with phosphorylation, consistent with crystal structures of NBD1-RE that included R region residues from 654–673, the regulatory extension. These show residues 655–668 forming a helix in the nonphosphorylated ATP-bound form and the helix extending only from residues 658–668 in the phosphorylated state (Lewis, Buchanan et al. 2004).

Figure 2.5: Free R region SSP values

SSP values calculated for the free phosphorylated or nonphosphorylated R region and averaged over a sliding window of 5 residues. Positive SSP values reflect the fractional α-helical structure for each residue in a weighted average of R region conformations. Negative SSP values reflect the fractional β-strand structure. Red circles indicate phosphorylation sites.

2.4 Discussion

The R region regulates CFTR in a complex manner with a dependency on multiple PKA sites and sensitivity to a variety of interaction partners. In order to aid in understanding the molecular basis for this mode of regulation we must begin by probing the structural properties of the isolated R region. The isolated R region is globally disordered, meaning that it samples multiple, heterogeneous conformations having varying degrees of compactness and secondary structural features with rapid inter-conversion between conformers (Figure 2.3). To enable site-specific structural characterization, we assigned the resonances in the spectrum to specific nuclei. We were able to assign
97% and 99% respectively of the nonphosphorylated and phosphorylated R region backbone resonances (Figure 2.4), one of the longest disordered protein sequences assigned to date. Distinct segments of the R region have fractionally populated local helical structure, reflecting a bias towards helical conformations for these residues in the pool of conformers (Figure 2.5). These helices fluctuate within the ensemble, with some conformers having helices that span over 20 residues, while other conformers only have helical conformations for residues with the highest fractional secondary structure. Phosphorylation reduces the bias towards helical conformations, except in the vicinity of Ser768. The decrease in helicity seen N-terminal to Ser737 is consistent with the observation that phosphorylation at this site leads to a marked gel shift due to a conformational change (Csanady, Seto-Young et al. 2005), which may also reflect changes in non-local structure.

The modulation of secondary structural propensity with phosphorylation through the altered stability of helices provides structural insight into a potential mechanism for the phospho-regulation of R region interactions via altered R region properties. In addition, the availability of nonphosphorylated and phosphorylated R region spectral assignments enables the mapping of structural probes in other NMR experiments to the R region sequence. With the spectral assignments as a basis, further experiments will provide an even more complete characterization of the R region free state to inform our understanding of R region interactions.
Chapter 3: Structural Characterization of the Free R Region

I performed the NMR experiments and analyzed the NMR data presented in this chapter. SAXS scattering experiments on nonphosphorylated R region were performed by Alexander Grishaev at the NIH. The relaxation data presented in this chapter is published in Jennifer M.R. Baker, Rhea P. Hudson, Voula Kanelis, Wing-Yiu Choy, Patrick H. Thibodeau, Philip J. Thomas, and Julie D. Forman-Kay. CFTR Regulatory (R) Region Interacts with NBD1 Predominantly Via Multiple Transient Helices. (2007) Nature Structural and Molecular Biology, 14, 738-745. The RDC values presented in this chapter are published in Joseph A. Marsh, Jennifer M.R. Baker, Martin Tollinger, and Julie D. Forman-Kay. Calculation of residual dipolar couplings from disordered state ensembles using local alignment. (2008) Journal of the American Chemical Society, 130, 7804-7805. The method presented in this paper was used in the program ENSEMBLE to analyze RDC values. ENSEMBLE calculations were performed by Joseph A. Marsh and I analyzed the calculated ensemble of conformations. Results in this chapter, besides the relaxation data and RDC values, will be included in a larger paper comparing calculated ensembles of the nonphosphorylated and phosphorylated R region.
3.1 Introduction

Despite the central role the R region plays in CFTR regulation, its characterization has been very limited at the structural level. Having completed spectral assignments of the R region (Chapter 2), we built upon this data using NMR techniques, as well as small angle x-ray scattering (SAXS), to probe the structural properties of the R region, focusing primarily on the nonphosphorylated state. Transient 2° and 3° intra-R region contacts were probed using fast timescale relaxation (motional properties), RDCs (bond vector angles with respect to a common reference frame), and PREs (inter-residue distances). Additionally, global properties of the R region conformational ensemble were examined by measuring the translational diffusion coefficient, giving the hydrodynamic radius, and the SAXS scattering curve. With information from this variety of structural probes, as well as chemical shift values (Chapter 2), we used the program ENSEMBLE to calculate a representative conformational ensemble of the nonphosphorylated R region free state. Having examined the wild-type R region, structural consequences of the CF-causing mutations F693L, V754M, and D806G located in regions outside of the PKA consensus sequences were also examined.

3.2 Experimental procedures

3.2.1 Protein preparation

Human CFTR (P13569) R region was expressed and purified as in section 2.2.1 with uniform $^{15}$N/$^{13}$C isotopic labeling in BL21(DE3) CodonPlus E. coli grown in M9 media, but without HPLC reverse phase purification. Selected R region samples were phosphorylated as in section 2.2.2. For paramagnetic relaxation enhancement (PRE) experiments (section 3.2.6), a series of R region single cysteine mutants were generated using QuikChange mutagenesis (Stratagene) including S678C, N706C, L732C, S756C, and A789C in a C832S background with the single wild-type R region cysteine removed. CF-causing mutations (http://www.genet.sickkids.on.ca/cftr/app) were introduced into selected single-cysteine mutant backgrounds including F693L (S678C), V754M (L732C), and D806G (A789C). All R region missense mutants were purified as for wild-type.
3.2.2 Nanosecond-picosecond timescale relaxation experiments

Uniformly $^{15}$N/$^{13}$C labeled nonphosphorylated and phosphorylated R region was prepared (3.2.1). $^{15}$N $R_I$, $R_{1p}$, and heteronuclear NOE values were measured using previously published pulse schemes (Farrow, Muhandiram et al. 1994) on 0.25 mM R region samples (in 125 mM K$^+$ phosphate (pH 6.8), 125 mM KCl, 2 mM EDTA, 2 mM benzamidine, 2 mM DTT). $^{15}$N $R_I$ values were measured from 2D spectra recorded with the relaxation delays 10.1, 70.6, 151.2, 252.9, 362.9, 493.9, and 655.2 ms. $^{15}$N $R_{1p}$ values were measured from 2D spectra recorded with delays of 2, 13, 28, 45, 66, 90, and 120 ms. $^{15}$N $R_2$ values for each residue were calculated by correction of the observed relaxation rate $R_{1p}$ for the offset of the applied spin-lock rf field to the resonance using the equation $R_{1p} = R_I \cos^2 \theta + R_2 \sin^2 \theta$, where $\theta = \arctan(\omega_{SL}/\Delta \omega)$. $\omega_{SL}$ was 1824.8 Hz and 1811.6 Hz for nonphosphorylated and phosphorylated R region respectively. Steady-state NOE values for both nonphosphorylated and phosphorylated R region were obtained with saturation of $^1$H for 5 s and a 10 s delay between scans and without $^1$H saturation using a 15 s delay between scans. All data sets were processed using the program "Fuda: A function and data fitting and analysis package" (personal communication, S.M. Kristensen, Chemistry Department, University of Copenhagen, and D.F. Hansen, Department of Medical Genetics, University of Toronto).

3.2.3 Hydrodynamic radius experiments

Uniformly $^{15}$N/$^{13}$C labeled nonphosphorylated and phosphorylated R region samples at 0.25 mM were prepared (3.2.1) and exchanged into 125 mM K$^+$ phosphate (pH 6.8), 125 mM KCl, 2 mM DTT with 0.04% dioxane as an internal diffusion rate standard. Pulse field gradient NMR diffusion experiments were recorded using the PG-SLED pulse sequence (Wilkins, Grimshaw et al. 1999) on a Varian Inova 500 MHz spectrometer with a room temperature probe and actively shielded z-gradients at 10 °C. Encoding and decoding gradients were used at 4, 8, 12, 16, 20, 24, 28, 40, 44, 48, 52, 56, and 60 G/cm, with the experiments recorded in triplicate. Protein amide peaks from 8.0 - 9.0 ppm were integrated for both nonphosphorylated and phosphorylated R region samples. A dioxane peak was integrated over 3.7 - 3.83 ppm (3.74 - 3.82 ppm) in the nonphosphorylated (phosphorylated) R region samples. The integrated peak intensities
were fit to a single exponential decay curve using the routine dfitlinux in the program Dasha (Orekhov, Nolde et al. 1995). The hydrodynamic radius was determined by comparing protein diffusion rates to diffusion rates for the internal standard dioxane.

3.2.4 Small-angle x-ray scattering

Solution scattering data were acquired by Alexander Grishaev at the NIH on a 6 mg/mL nonphosphorylated R region sample in 125 mM K$^+$ phosphate (pH 6.8), 125 mM KCl, 2 mM EDTA, 2 mM benzamidine, 2 mM DTT using a SAXSess instrument (Anton-Paar). Data were collected as series of sequential 1 hr acquisitions on the sample, followed immediately by the matching dialysate buffer. The scattering curves were processed using GNOM software (Svergun 1992) to account for the geometry of the experimental setup.

3.2.5 Residual dipolar couplings

HNCO-based, non-TROSY 3D experiments (Goto, Skrynnikov et al. 2001) were recorded at 10 °C in the presence and absence of Pf1 phage (deuterium splitting of 24 Hz) to measure $^1$H-$^{15}$N RDCs (Clore, Starich et al. 1998; Hansen, Mueller et al. 1998). Experiments were performed on a Varian Inova 600 MHz spectrometer with a cold probe and actively shielded gradients. Data sets were processed with NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed with PIPP (Garrett, Powers et al. 1991). For most residues, anisotropic conditions gave rise to larger splittings (J+D) than isotropic conditions (J only). The sign convention of subtracting (J) from (J+D) was used, yielding mostly positive values of the dipolar coupling.

3.2.6 Paramagnetic relaxation enhancement experiments

Uniformly $^{15}$N/$^{13}$C labeled wild-type and single-cysteine mutant samples were prepared (3.2.1). Samples were dialyzed against degassed, argon purged 6 M GdmCl, 20 mM tris (pH 8), 200 mM NaCl with 5 mM fresh DTT to ensure samples were reduced before exchanging into DTT-free buffer using a PD-10 column (Pharmacia). S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) (Toronto Research Chemicals) was immediately added from an acetonitrile stock in a >10 molar excess compared to protein. The MTSL coupling reaction proceeded at 4 °C overnight.
in the dark. Excess MTSL was removed by dialyzing against 6 M GdmCl, 20 mM tris (pH 8), 200 mM NaCl, then samples were renatured by dialyzing against 125 mM K+ phosphate (pH 6.8), 125 mM KCl. HNCO experiments (Sattler, Schleucher et al. 1999) were recorded with oxidized spin label, samples were reduced with a 5-fold molar excess of ascorbate at room temperature for at least 1 hour, and experiments were repeated. Experiments were recorded with 8 transients (200 µM L732C, 280 µM A789C, 70 µM A789C/D806G), 16 transients (80 µM S678C, 100 µM N706C), or 16 transients in the oxidized state and 8 transients in the reduced state (100 µM S756C) on an 800MHz Varian Inova spectrometer at 10 °C with a room temperature, triple-resonance probe with actively shielded gradients. Experiments were recorded with 8 transients (150 µM C832) on a 600MHz Varian Inova spectrometer at 10 °C with a cold, triple-resonance probe with actively shielded gradients. Peak intensities were quantified using PIPP (Garrett, Powers et al. 1991) and errors quantified by propagation of the signal-to-noise measured in NMRDraw (Delaglio, Grzesiek et al. 1995). A789C spectra indicated a slight incomplete labeling of the protein; to compensate, 0.2 was subtracted from peak intensity ratios, a value chosen to be in agreement with D806G(A789C) ratios for residues distant from the D806G mutation.

3.2.7 ENSEMBLE calculations

Various structural probes of the nonphosphorylated R region were used as inputs into the program ENSEMBLE (Choy and Forman-Kay 2001; Marsh, Neale et al. 2007; Marsh and Forman-Kay 2008) to calculate a structural representation of the R region free state. The SAXS scattering curve was used directly as a restraint in calculations. An initial pool of 1000 conformers was generated using TraDES (Feldman and Hogue 2000), with 250 conformers from each of four different secondary structure distributions: ‘GOR3’, ‘Coil’, ‘Alpha’, and ‘Beta’, where the latter two are set to have 60% of either alpha or beta structure, respectively. From this pool, 100 structures were randomly selected and the energy of the conformational ensemble compared to the experimental restraints using Monte Carlo sampling, and evaluated using the Metropolis criterion. This process was repeated, with additional structures continually added to the pool by modifying previously selected conformers using: (1) the Unfoldtraj utility of TraDES with a random temperature between 500-1300 and a random timestep of 40-140 or (2) by
using the CNS (Brunger, Adams et al. 1998) input file ‘anneal.inp’ to add secondary structure by introducing α-helical or β-strand dihedral angle restraints for a randomly chosen 4-12 residue segment and hydrogen bond restraints for adding α-helix.

Additional random structures were also added to the pool using TraDES (Feldman and Hogue 2000). Once 100 structures could be selected from the pool to satisfy the experimental restraints, subsequent calculations attempted to satisfy the restraints with a decreasing number of conformers until an equilibrium was reached with no reductions in conformer pool size for 3 or more days (as run on the High Performance Facility at the Hospital for Sick Children Centre for Computational Biology). Three sets of calculations beginning from the minimized pool of 100 structures were completed to generate three ensembles describing the R region free state.

3.2.8 CF-causing R region mutant chemical shifts

HNCO experiments (Sattler, Schleucher et al. 1999) were recorded on CF-causing (in context of a single cysteine) R region mutants: 280 µM F693L(S678C), 240 µM V754M(L732C), and 100 µM D806G(A789C), with 8 transients on a Varian Inova 800MHz spectrometer at 10 °C with a room temperature, triple resonance probe with actively shielded gradients. Buffer conditions were 125 mM K⁺ phosphate (pH 6.8), 125 mM KCl, 2 mM EDTA, 2 mM benzamidine, 2 mM DTT. An HNCACB experiment (Sattler, Schleucher et al. 1999) was also recorded with 8 transients on 280 µM F693L(S678C) to assign ¹H, ¹⁵N, ¹³C, ¹³Cα, and ¹³Cβ resonances corresponding to residues near the C832S mutation (which removes the naturally-occurring cysteine). All spectra were referenced using the internal reference DSS (sodium 2,2-dimethyl-2-silpentace-5-sulfonate). Data was processed using NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed using NMRView (Johnson and Blevins 1994).

3.3 Results

3.3.1 Phosphorylation reduces R region structural contacts

Residues involved in structural contacts within the isolated R region were probed by examining their dynamic properties. Motions on fast timescales were quantified using $R_1$, $R_{1p}$, and heteronuclear NOE relaxation experiments (Farrow, Muhandiram et al.
1994) (Figure 3.1), with $R_2$ relaxation rates calculated from $R_1$ and $R_{1\rho}$ relaxation rates (3.2.2), plotted in Figure 3.1a-b. Rates could not be calculated for some residues because of the substantial spectral overlap, as the experiments were recorded as two dimensional $^1\text{H}^n$-$^{15}\text{N}$ correlations. Low $R_2$ relaxation rates reflect rapid motion (on the ps-ns timescale) such as conformer interconversion for disordered proteins, while higher relaxation rates are indicative of restricted mobility or slower fluctuations between conformational states (on the µs-ms timescale). In both nonphosphorylated and phosphorylated states the R region has $R_2$ relaxation rates generally in the range of 4–8 s$^{-1}$, consistent with the values expected for a disordered protein and lower than those expected for a folded protein with the same number of residues. Uniformly low $R_2$ relaxation rates throughout the sequence with further reduced rates at the termini are expected for a model disordered protein with no structural contacts (Schwalbe, Fiebig et al. 1997). In contrast, the $R_2$ relaxation rates for the R region are not uniform throughout the primary sequence, with higher rates observed for a number of residues including His667, Glu733, Ser768, and Ser809 that are within regions having >5% helical population and close to phosphorylation sites. This suggests that the restriction of motion by secondary structural contacts contributes to higher $R_2$ relaxation rates, with potential additional contributions from the exchange of individual conformers between different conformational states, tertiary contacts, or hydrophobic clusters (Schwalbe, Fiebig et al. 1997). Phosphorylation generally decreases $R_2$ relaxation rates, including near residues 680, 735, 810, and 835, while rates near 770 are increased (Figure 3.1b).

$R_1$ relaxation rates are relatively uniform throughout both nonphosphorylated and phosphorylated R region (Figure 2.1c-d). The heteronuclear NOE values (Figure 2.1e-f) are positive, with the exception of S654 at the phosphorylated R region N-terminus, and vary throughout the sequence of both nonphosphorylated and phosphorylated R region. Positive heteronuclear NOE values for the R region indicate that motion on the picosecond-nanosecond timescale is somewhat restricted compared to the negative values expected for highly mobile residues (Kay, Torchia et al. 1989). Slightly increased heteronuclear NOE values occur in similar residue segments to where $R_2$ relaxation rates are increased, another indication of restricted motion for these residues. The NOE values are modestly lower for phosphorylated R region than for
nonphosphorylated R region, although the error bars are larger in the nonphosphorylated dataset because of greater overlap and decreased signal-to-noise in these spectra.

![Figure 3.1](image)

**Figure 3.1** Fast timescale motion of the free R region

$R_2$ relaxation rates are shown for (a) nonphosphorylated and (b) phosphorylated R region. $R_1$ relaxation rates are shown for (c) nonphosphorylated and (d) phosphorylated R region. Heteronuclear NOEs are shown for (e) nonphosphorylated and (f) phosphorylated R region. Error bars in (a-d) indicate propagation of errors calculated from the covariance matrix of the least-squares fit of $R_1$ and $R_{1p}$ values. Error bars in (e-f) indicate propagation of errors from the signal-to-noise ratio. Red circles indicate PKA phosphorylation sites.

### 3.3.2 Molecular size distribution

Molecular dimensions of the R region conformer pool were probed using two complementary techniques, NMR pulse field gradient (PFG) measurements and SAXS.
PFG measurements were used to calculate the hydrodynamic radius ($R_h$). Average diffusion coefficients ($D_t$) in the nonphosphorylated R region sample were fit as $8.6(\pm 0.1) \times 10^{-10}$ cm$^2$ s$^{-1}$ (protein) and $1.25(\pm 0.02) \times 10^{-8}$ cm$^2$ s$^{-1}$ (dioxane), giving an $R_h$ value of $31(\pm 1)$ Å (Wilkins, Grimshaw et al. 1999). In the phosphorylated R region sample, $D_t$ values were $1.09(\pm 0.01) \times 10^{-9}$ cm$^2$ s$^{-1}$ (protein) and $1.39(\pm 0.01) \times 10^{-8}$ cm$^2$ s$^{-1}$ (dioxane), giving an $R_h$ value of $27.2(\pm 0.9)$ Å, and indicating that the phosphorylated R region conformational ensemble is slightly more compact. These values were recalibrated using an $R_h$ value for the lysozyme reference calculated using HYDROPRO (Garcia De La Torre, Huertas et al. 2000), as this method was also used to calculate $R_h$ for conformers in ENSEMBLE. This yielded $R_h$ values of $32.6 (\pm 1.0)$ Å (nonphosphorylated R region) and $28.6 (\pm 0.9)$ Å (phosphorylated R region). SAXS data was also collected for the nonphosphorylated R region and the scattering curve used as an input into ENSEMBLE (see 3.2.7).

3.3.3 R region bond angle orientation

To further characterize the nonphosphorylated R region, $^{15}$N-$^1$H bond angle orientations were probed by recording RDCs (Figure 3.2) in Pf1 phage. The values are highly non-uniform and range from 6.7 to -10.7 Hz. Negative RDC values occur for residues 654-662, 680, 760-763, 769-770, and 803-810, segments with the largest α-helical propensities (Figure 2.5, yellow bars in Figure 3.2). The most positive RDC values, at residues 709, 750, and 824 (except 786), are mostly in segments with β-strand propensities (Figure 2.5, blue bars in Figure 3.2). This loose correlation between sign of RDC value and SSP value is similar to behaviour noted in acid-denatured ACBP (Fieber, Kristjansdottir et al. 2004) and acid-denatured apomyoglobin (Mohana-Borges, Goto et al. 2004). The absolute sign of the RDC values is opposite in the case of R region because of the use of different alignment media (phage, R region versus stretched polyacrylamide gels, ACBP and apomyoglobin). While folded proteins are thought to have a relatively uniform global alignment with respect to the media, the conformational heterogeneity of disordered proteins merits a more complex treatment where each conformer is considered individually with respect to the media. One interpretation of RDCs in disordered or denatured proteins is that they originate from transient alignment of protein segments with α-helical or β-strand propensity (Mohana-
Borges, Goto et al. 2004; Marsh, Baker et al. 2008). The N-H bond vectors in a β-strand segment align perpendicular to the chain direction (and perpendicular to the principle axis of the local alignment tensor), yielding positive RDC values in phage alignment medium. In contrast, N-H bond vectors in an α-helical segment align parallel to the chain direction (and the principle axis of the local alignment tensor), so that the RDCs have negative values (Mohana-Borges, Goto et al. 2004).

![Figure 3.2](image)

**Figure 3.2** Residual dipolar couplings for the nonphosphorylated R region

RDC values, determined by subtracting the J coupling from the D+J (dipolar and J) coupling. A reference dashed red line indicates a coupling value of zero. Shaded bars above the plot indicate segments where at least three consecutive residues have greater than 5% α-helical (yellow) or β-strand (blue) propensity (Figure 2.5).

### 3.3.4 Transient tertiary contacts throughout the nonphosphorylated R region

Transient medium to long-range contacts throughout the nonphosphorylated R region primary sequence were probed using paramagnetic relaxation enhancement experiments. Single cysteines, introduced at 5 positions throughout the primary sequence and coupled with site-specific MTSL spin probes, caused few chemical shift changes from wild-type outside of the vicinity of the mutation. Minor changes are observed for residues ~715-730, flanking residues 718-722 that could not be assigned because of µs–ms timescale broadening, indicating that this is a more structurally-sensitive segment of the R region. Overall, the mutations and spin label coupling cause limited perturbations to the 2° and 3° structure in the molecule (Figure 3.3).
Figure 3.3  R region single cysteine mutant chemical shift changes

Total chemical shift change for $^{1}H_{N}$, $^{15}N$, and $^{13}C'$ nuclei, weighted by gyromagnetic ratio, from HNCO experiments. Wild-type and single cysteine mutant R region with MTSL coupled (green triangle) are compared for a) S678C, b) N706C, c) L732C, d) S756C, and e) A789C, all in a C832S background to remove the native cysteine. A dashed red reference line indicates the average total chemical shift change for all residues more than 10 away in primary sequence from the respective mutation site, calculated for each mutant. Minor chemical shift referencing offset adjustments were made by subtracting the average change from wild-type over a 40 residue window: residues 780-820 (S678C, L732C, S756C) or residues 730-770 (N706C, A789C). No chemical shifts were available for C832S R region, so this same calculation was applied to residues 828-838, except with S678C/C832S chosen as the reference state.
Wild-type (cysteine at residue 832) and 5 mutants were examined with the spin label in oxidized and reduced states to probe through-space contacts with other residues (Figure 3.4). In all cases, many resonances corresponding to residues distant in the primary sequence were affected by the spin probe, indicating that the nonphosphorylated R region has many transient medium to long-range contacts in the free state. The observed resonance broadening could arise from either moderate proximity of specific pairs of residues in a large fraction of conformers or tighter contacts formed in a smaller fraction of conformers. A long stretch of residues from 657 to 722 are significantly affected ($I_{\text{ox}}/I_{\text{red}} < 0.5$) by a spin probe at residue 678, indicating the presence of multiple contacts within the N-terminal region (Figure 3.4a). A spin probe at residue 706 causes broadening throughout the sequence, indicating that this residue is in transient proximity to most residues (Figure 3.4b). Resonances corresponding to residues 718-722 could not be assigned in nonphosphorylated wild-type R region, likely because of µs-ms timescale motion. The large influence of a spin probe at 706 on peak intensities for residues throughout the sequence may arise because of their contacts with residues 718-722. Residues 707-772 throughout the centre of the R region sequence are significantly broadened ($I_{\text{ox}}/I_{\text{red}} < 0.5$), (except $I_{\text{ox}}/I_{\text{red}}=0.53$ for E743) by the presence of a spin label at position 732, indicating the presence of additional significant transient contacts throughout the centre of the sequence (Figure 3.4c). Spin labels at residues 756 and 789 (Figure 3.4d-e) have the greatest effect local to the spin label and smaller effects more distant in the sequence where $I_{\text{ox}}/I_{\text{red}}$ is between 0.5 and 1.0. The R region C-terminus, probed with a spin label at 832, has moderate contacts throughout the C-terminal half of the protein, with most residues from 751-838 having $I_{\text{ox}}/I_{\text{red}}$ ratios of 0.6 or less. Taken together, these experiments indicate the presence of transient contacts throughout the molecule. Stretches of approximately 65 residues near the N-terminus and centre of the molecule, and approximately 85 residues near the C-terminus are involved in these transient contact clusters.
Figure 3.4 Transient secondary and tertiary contacts throughout the R region

Peak intensity ratios are calculated between HNCO spectra with covalently coupled MTSL in the oxidized (active) state ($I_{ox}$) to the reduced (inactive) state ($I_{red}$). MTSL is coupled at residues (green triangle) (a) 678, (b) 706, (c) 732, (d) 756, (e) 789, and (f) 832. A dashed red line in (a) illustrates the approximate behaviour expected for a protein with no secondary or tertiary contacts.
3.3.5 Hydrodynamic and $2\theta$ structure properties of calculated ensembles

With structural information from a variety of experimental probes of the nonphosphorylated R region available, the program ENSEMBLE (Marsh, Neale et al. 2007; Marsh and Forman-Kay 2008) was used to generate a structural representation of the nonphosphorylated R region free state. Calculations were carried out in triplicate, and minimized to a similar number of conformers, 30, 32, and 32 respectively. These conformers have a variety of $2\theta$ and $3\theta$ structural features and sample a range of hydrodynamic properties. There is a broad range of $R_h$ and $R_g$ values for the conformers (Figure 3.5), from 26-44 Å ($R_h$) and 21-33 Å ($R_g$), with a bimodal distribution. This is compared to the narrow, unimodal distribution expected for a folded protein or the broad, unimodal distribution expected for a completely unfolded protein. During calculations, these two parameters tended to pull the conformer pool in opposite directions, towards more compact conformers ($R_h$) or more expanded conformers (SAXS scattering). This may reflect inherent limitations of both types of measurement either through the presence of minor aggregation (SAXS) or approximations such as the dependence of $R_h$ on protein chain length. Also, these two parameters are most sensitive to conformers with size properties at opposite ends of the distribution; $R_h$ is most sensitive to compact conformers, while SAXS scattering is most sensitive to extended conformers (Choy, Mulder et al. 2002). The ensemble average $R_h$ was 35.1 Å, reasonably close to the measured value of 32.6 Å. The SAXS scattering curve overlays well with the experimental data (Figure 3.5c). With their sensitivity to different properties, including both $R_g$ and $R_h$ parameters in ENSEMBLE calculations enables the best description of the ensemble of conformations present.
Figure 3.5 Ensemble hydrodynamic properties

The distribution of (a) $R_h$ and (b) $R_g$ are plotted. Error bars correspond to the standard deviation between 3 sets of calculations. (c) Comparison of SAXS scattering curves between experimental (magenta) and conformer (blue) values.

The RDC values calculated from the ensemble by predicting the alignment of an overlapping 15-residue sliding window (Marsh, Baker et al. 2008) are also in very good agreement with experimental values, with experimental values falling within the standard deviation of ensemble runs (Figure 3.6a). Averaged conformer sampling of Ramachandran space is compared to experimental SSP values (Marsh, Singh et al. 2006) in Figure 3.6b-d. There is generally good agreement between Ramachandran space sampling (Figure 3.6b,c) and SSP values (Figure 3.6d). Minor differences occur for residues 741-746 and 783-790, which have 5-28% net fractional strand and 5-19% net fractional helix by SSP values, respectively, but sample approximately 50% alpha and 55% beta space, respectively, in the ensemble calculations.
Figure 3.6 Comparison of ensemble and experimental values

(a) Comparison of RDC values between experimental values (magenta) and calculated ENSEMBLE conformer values (blue). Error bars represent the standard deviation between three different ENSEMBLE calculations. (b) Fractional ENSEMBLE conformer sampling of alpha (red) and beta (blue) Ramachandran space is plotted. Values are averaged over all R region conformers from three R region ENSEMBLE calculations and dotted lines indicate the standard deviation between the three calculated ensembles. (c) Difference between alpha sampling and beta sampling from (b). A red reference line indicates the zero value. Asterisks indicate regions having significant differences between Ramachandran space sampling and SSP values. (d) Experimental SSP values calculated from Cα and Cβ chemical shifts.
3.3.6 R region conformational ensemble

To examine R region conformers within the heterogeneous ensemble, the 94 conformers from the 3 R region ENSEMBLE runs were initially pooled together, and then clustered based on the Cα-Cα distance matrix RMSD (Choy and Forman-Kay 2001; Marsh and Forman-Kay 2008). Conformers clustered into 10 groups (Figure 3.7), displayed in ascending order by $R_g$. The random conformer displayed from each cluster readily illustrates the conformational heterogeneity present. Several major patterns of residue contacts are present in the various clusters: overall compaction (A, B, C), longer-range contacts involving the conformers bending back on themselves (E, F, I), and clustering at the N- and/or C-termini (D, G, H, J). Long-range contacts involving residues near the centre of the protein are less common because they impose more entropic penalty than those closer to the termini. The presence of more long-range contacts near the termini (illustrated in cluster H with a green dashed line) has been previously noted in polymer simulations (Chan and Dill 1989). Tighter long-range contacts are found in several clusters including B (~790 with ~654-660 and 680, C (790-800 with 685-690), E (~790 with 690-715), J (790-810 with 750-765), J(825-838 with 740-760), and J(830-838 with 800-810), illustrated with blue dashed ellipses below the diagonal (Figure 3.7). Each of these clusters with tighter contacts represents a small fraction of the total conformers (0.03-0.06); however these conformers are still necessary to satisfy the experimental data and must contain important structural information.
Figure 3.7 Clustered R region conformers

Conformers from the three R region ENSEMBLE calculations were clustered based on $C\alpha-C\alpha$ distances. The resulting 10 clusters (A-J) are shown, ordered from the most compact to most extended cluster. Residue-residue distance contact plots, the fraction of conformers in each cluster ($f$), average $R_g$, and a sample conformer are shown for each cluster. Blue dashed ellipses indicate residues with long-range contacts. Green dashed arcs illustrate the preference for more long-range contacts near the N- and C-termini.

To examine all the transient contacts present within the R region ensemble, a pairwise distance contact plot is shown for residue-residue pairs in Figure 3.8. Overall, the transient contacts throughout the R region are very heterogeneous. The location of spin labels for PRE measurements could have introduced bias towards contacts for these residues because they have more structural restraints than other residues in the calculations. There appears to be very limited bias (note locations of dashed lines in Figure 3.8), except potentially near the 789 spin label site. However, residue S790 has been implicated in free R region contacts that are disrupted upon NBD1 binding (see 4.3.1), so these likely represent real contacts (near S790 with residues 675-745, blue dashed ellipse). R region residues with greater than expected contacts (blue shading) occur for the most N-terminal residues to ~660 with residues near 680. The first residues of the R region have helical propensity (Figure 2.5) and form a transient amphipathic helix. The only tryptophan residue in the R region is Trp679, so transient hydrophobic clustering may occur between these segments. Other greater than expected interactions occur between residues ~785-810 and 820 to the C-terminus that may be mediated by charge. Most negatively charged residues within the R region occur in discrete segments involving residues 725-734, 741-746, and 815-838 (Figure 3.9a). Medium- to long-range interactions involving residues 725-734 or 741-746 would be disfavoured by a relatively large loss of conformational entropy because these residues are close to the centre of the R region sequence. Conversely, residues 815-838 have a very high concentration of negative charge (residues 822, 823, 826, 827, 828, 831, 835, 836, 838) and are near the C-terminus so contacts involving these residues would have a lower entropic penalty. Interactions with the C-terminus could involve dibasic motifs at residues 785/786, 792/793, and 810/811.
Figure 3.8 R region conformational ensemble contact map

Average residue-residue C$_{\alpha}$-C$_{\alpha}$ distances across all R region conformers, with the diagonal representing self-contact distances. Values are normalized based on the average distance for all residues a given distance apart in primary sequence. Residues that are closer in space than average are shaded blue; those further away in space are shaded red. Dashed black lines indicate the positions of PRE spin labels in the input data. The dashed blue ellipse indicates enriched contacts involving residues near S790.

Shorter range contacts, approximately 10-20 residues off the diagonal, loosely follow the R region hydrophobicity profile (Figure 3.9b). Many R region segments with higher than average hydrophobicity (Figure 3.9b, above red dashed line), including approximately residues 660-680, 705-715, 735-740, 745-750, and 770-775, have more contacts, while residues 725-730 and 785-795 with lower than average hydrophobicity have fewer contacts.
Figure 3.9  R region sequence properties

(a) R region residues 654 to 838 are shown, with positively (blue) and negatively (red) charged residues indicated. (b) Kyte-Doolittle Hydrophobicity (Kyte and Doolittle 1982) calculated using a 5 residue window in ProtScale (Gasteiger, Hoogland et al. 2005).

3.3.7 Characterizing R region CF-causing mutations

The majority of CF-causing missense mutations in the R region are located in the PKA consensus sequences and could cause disease by altering phosphorylation levels, leading to altered CFTR trafficking or activity. However, there are three candidate mutations F693L, V754M, and D806G that cause disease, yet are not located within these consensus sequences (http://www.genet.sickkids.on.ca/cftr/app). To determine if these mutations alter structural properties of the R region free state, as a potential contribution to the disease-causing mechanism, NMR data for these mutants were compared to the wild-type R region. The $^1$H, $^{15}$N, and $^{13}$C chemical shifts of these mutants were assigned by comparing mutant spectra to those of wild-type R region. Chemical shifts are sensitive to local 2° and 3° contacts of a given nuclei, so structural changes introduced by the mutations will result in chemical shift changes, quantified for F693L, V754M, and D806G (Figure 3.10).
Figure 3.10 Chemical shift changes with CF-causing mutations

Total chemical shift change, weighted by gyromagnetic ratio, for $^{1}H$, $^{15}N$, and $^{13}C'$ nuclei from HNCO experiments. CF-causing R region mutants (red star) are compared to their background single cysteine-MTSL coupled mutants (green triangle) for (a) F693L (in context of S678C), (b) V754M (in context of L732C), and (c) D806G (in context of A789C). Chemical shift changes for F693L residues F693, G694, E695 (a) are truncated (indicated by ~); their actual values are 257, 258, and 379 Hz, respectively. A red dashed line indicates the average chemical shift change calculated for all residues over 10 away in primary sequence from the mutation site. Minor chemical shift referencing offset adjustments were made by subtracting average changes from the background single-cysteine mutants coupled to inactivated MTSL over residues 790-830 (S678C/F693L versus S678C), 660-700 (L732C/V754M versus L732C), and 730-760 (A789C/D806G versus A789C). Note that an HNCACB experiment (Sattler, Schleucher et al. 1999) was recorded on S678C/F693L, enabling the assignment of residues at the F693L mutation site (a).
In Figure 3.10, F693L initially appears to cause very large chemical shift changes at the mutation site (red star) compared to V754M and D806G. This is because an additional HNCACB spectral assignment experiment was performed for this mutant (3.2.8), allowing the assignment of resonances right at the F693L mutation site. In V754M and D806G many chemical shift changes close to the mutation site could not be followed, and these bars are missing in Figure 3.10b,c. In D806G resonances corresponding to residues up to 9 C-terminal to the mutation site could not be followed. These CF-causing mutations cause similar chemical shift changes over residues ~715-717 and 723-730, flanking residues 718-722 that had μs-ms timescale broadening and could not be assigned, as the single cysteine mutations (Figure 3.3), indicating the structural sensitivity of this segment. The V754M mutation causes few other significant changes, with modest chemical shift changes near 765-775, a segment with some of the highest \( R_2 \) relaxation rates (Figure 3.1), that are also noted in F693L and D806G. The F693L and D806G mutations cause more chemical shift changes, indicating that these mutations have more severe structural consequences than V754M. In F693L, residues from 675-699 have significant chemical shift changes, up to 18 residues away from the mutation site. The D806G mutation causes chemical shift changes in clusters of residues including near 660, 670, 690, and 820, indicating that transient 2° and 3° contacts are altered at additional sites throughout the molecule. This mutation introduces a glycine residue into an R region segment with α–helical propensity (Figure 2.5), likely disrupting this transient helix and altering contacts throughout the R region.

To further probe altered contacts in D806G R region, PRE experiments were compared to those for the A789C reference (Figure 3.11). Despite the D806G mutation, transient contacts are similar throughout much of the molecule with the exception of residues 771-773 having greater contacts with residue 789 (site of the spin label) than in wild-type. This, in combination with chemical shift changes from the mutation (Figure 3.10c), indicates that structural changes propagated from D806G to the rest of the molecule, especially residues 765-770 and 775-779, must also change transient contacts between 771-773 and 789.
Figure 3.11 Altered transient contacts with the D806G CF-causing mutation

Peak intensity ratios are calculated between HNCO spectra with covalently coupled MTSL in the oxidized (active) state \( I_{\text{ox}} \) to the reduced (inactive) state \( I_{\text{red}} \). MTSL is coupled at residue 789 (green triangle). (a) A789C/C832S wild-type reference (as in Figure 3.5e) and (b) D806G mutant (red star) in A789C/C832S background. The A789C \( I_{\text{ox}} \) peak intensities are scaled by a common factor derived from best fit between \( I_{\text{ox}}/I_{\text{red}} \) ratios for residues 654-750 common between both datasets to compensate for slight, unknown incomplete MTSL labeling of A789C. D806G residues with altered transient contacts are indicated (magenta asterisk).

3.4 Discussion

The R region is the regulatory centre of CFTR, critical to normal channel function by mediating interactions both within CFTR and with other interaction partners. There has been limited site-specific R region structural information to date; the studies presented here use a variety of structural probes to examine all R region residues. Chemical shifts, relaxation experiments \( (R_1, R_2, \text{ and heteronuclear NOE}) \) and \( R_h \) measurements have been used to characterize both the nonphosphorylated and phosphorylated R
region. These indicate that phosphorylation generally reduces structural contacts and leads to modest compaction. To further characterize the nonphosphorylated R region, the state most involved in NBD1 binding (see chapter 4), SAXS, RDCs, and PREs with a nitroxide spin label coupled at 6 different positions were also measured. Information from these various structural probes was integrated in the program ENSEMBLE (Marsh, Neale et al. 2007; Marsh and Forman-Kay 2008) to calculate a representation of the nonphosphorylated R region free state. For the first time, this enables a glimpse of individual conformers rather than the ensemble-averaged observables. ENSEMBLE calculations indicated that ~30 representative conformers were sufficient to explain the experimental data. This is a minimalist representation of the ensemble of R region conformers; sharp lineshapes observed in R region NMR spectra indicate that the R region is highly dynamic and samples many conformations. Each of the 30-32 conformers from these calculations is highly flexible and can be thought of as the ‘mean’ of a sub-ensemble of conformations that are accessible about this representative structure. The representative conformers are themselves highly heterogeneous and include contacts that could not be directly determined from the experimental datasets in isolation.

Various methods have been used to describe properties of denatured or intrinsically disordered proteins in solution. Two groups have used similar approaches to calculate RDC values (Bernado, Bertoncini et al. 2005; Jha, Colubri et al. 2005) or RDC values and SAXS scattering (Bernado, Blanchard et al. 2005) from an amino acid-specific statistical coil ensemble of conformations chosen from loop conformations in x-ray crystal structures outside of α–helices and β–strands. This method has been used to examine several proteins, each with thousands of conformations: 50000 for α–synuclein (Bernado, Bertoncini et al. 2005); 5000-16000 for Δ131Δ, chemically denatured apo-Myoglobin, and elginC, and ubiquitin (Jha, Colubri et al. 2005); 100000 for chemically denatured apo-Myoglobin, Δ131Δ, and nucleocapsid-binding domain of Sendai virus phosphoprotein (Bernado, Blanchard et al. 2005), and 50000 for tau protein (Mukrasch, Markwick et al. 2007). These calculated values can then be compared to experimental values, with differences between the two indicating potential secondary or tertiary contacts. In the case of α–synuclein, long-range tertiary contacts
introduced into the conformers used to calculate RDC values improved agreement with experiment values, consistent with the presence of these contacts in solution (Bernado, Bertoncini et al. 2005).

Other groups have used Monte Carlo or restrained molecular dynamics (MD) simulations, based on PRE distance restraints, to develop representations of GdmCl-denatured ACBP (Lindorff-Larsen, Kristjansdottir et al. 2004), acid-denatured ACBP (Kristjansdottir, Lindorff-Larsen et al. 2005), unfolded staphylococcal nuclease fragment Δ131Δ (Francis, Lindorff-Larsen et al. 2006) or intrinsically disordered α–synuclein (Dedmon, Lindorff-Larsen et al. 2005). These methods used multiple replicas, typically 20-25, to obtain reasonable $R_g$ values, which were simultaneously calculated such that they collectively satisfied the experimental restraints. Simulations were run for $2 \times 10^7$ Monte Carlo steps or 10-200ns (MD) after reaching equilibrium and each generated 4000-25000 conformers that were used for analysis. Residue-residue contact maps including all conformers for Δ131Δ and α–synuclein both indicated the presence of long-range contacts between residues >70 residues apart. These calculations used only PRE experimental data which has a $1/r^6$ dependence, making it relatively easy to satisfy restraints with specific, short-range contacts present in only a small number of conformers. This may be a limitation of an approach that uses only PRE experimental data.

The ENSEMBLE approach used in this work addresses these concerns and provides a more complete representation by incorporating a variety of types of structural information into the calculations. This includes the overall extended nature of the R region, reflected in its $R_h$ and SAXS scattering curves, its transient contacts based on $R_2$ relaxation rates, RDCs, and PREs, and its local conformations based on chemical shifts and RDCs. The ENSEMBLE calculations used in this work (Marsh and Forman-Kay 2008) represent an improvement over earlier versions (Choy and Forman-Kay 2001; Marsh, Neale et al. 2007) where conformers were selected from a pre-existing pool to produce agreement with experimental data. The improved methods used here (Marsh and Forman-Kay 2008) incorporate two new features: the ability to modify conformers that have been selected from an initial pool enabling constant generation of new conformers and addressing the problem of restricted sampling of conformational
space, and a minimization approach to determine the smallest number of conformers that can explain the data. After an initial equilibration period based on 100 conformers, the pool is reduced until the smallest possible number of conformers explains the experimental data. This approach addresses the over-fitting problem, recognizing the underdetermined nature of the system with a limited number of experimental restraints available in comparison to the large number of degrees of freedom.

The R region residue-residue contact plot (Figure 3.7) has a variety of transient contacts including residues near 790 that transiently interact with residues 675-750, contacts that are altered upon NBD1 binding. This cluster of contacts could be a key contribution to the overall degree of compactness of the R region. This contact plot is shown again to examine the secondary structural propensity of the free state where R region segments with 5% or more helical propensity (yellow bars) or more than 5% strand propensity (teal blue bars) are indicated on the axes (Figure 3.11). The R region segments with the highest helical propensities (Figure 2.5) also corresponding to segments with the highest $R_2$ relaxation rates (Figure 3.1a), are indicated with grey dashed lines at the boundaries. The behaviour of these segments is of particular interest because they mediate much of R region binding to NBD1 (see chapter 4). These transient helices are not more likely to be involved in transient interactions with other residues (areas between the dashed lines), with the exception of limited short to medium range interactions of 654-668 with those near 675, and residues 801-817 with those near the C-terminus. These helices also do not have significant transient interactions with each other. The R region binds multiple interaction partners (Naren, Cormet-Boyaka et al. 1999; Ko, Zeng et al. 2004), potentially also via these transient helices, so the relative accessibility of these helices in the free state is of interest. They may be relatively uninvolved in intra-R region interactions in the free state so that they are accessible and able to bind to a variety of interaction partners. Contacts involving other residues within the R region may be responsible for balancing the moderate compactness of the R region free state, with a potential role in localizing transient helices, while still permitting these transient helices to readily engage interaction partners.
Figure 3.12 Transient contacts of R region segments with \( \alpha \)-helical propensity

Average residue-residue \( \text{C}_{\alpha}-\text{C}_{\alpha} \) distances across all R region conformers, with the diagonal representing self-contact distances. Values are normalized based on the average distance for all residues a given distance apart in primary sequence. Residues that are closer in space than average are shaded blue; those further away in space are shaded red. Dashed lines delineate segments with the highest helical propensity and their contacts with other residues in the sequence.

Modulating this balance between accessibility and compactness within the R region free state may be important to maintain normal CFTR function. The disease-causing mutations F693L, V754M, and D806G involve R region residues outside the PKA consensus motifs. In particular, the D806G mutation introduces a Gly residue into a segment with 10-40% \( \alpha \)-helical propensity and causes chemical shift changes for residues throughout the R region (Figure 3.10c), indicating that it alters contacts.
Further examination of this mutation using PRE experiments (Figure 3.11) indicates that it strengthens transient contacts between residues 771-773 and 789 (the spin label site). The residues 771-773 lie within the sequence 760-783 that, when deleted, causes channel malfunction, underlying its importance (Balduresson, Ostedgaard et al. 2001). These residues also bind NBD1 (see 4.2.1), so strengthening interactions with other residues in the R region may hinder their ability to bind NBD1 and other partners, altering channel function.

Overall, a highly dynamic and heterogeneous representation of the R region emerges from this work, where the R region has the dual properties of overall extended molecular dimensions in comparison to a folded protein, while still possessing significant transient interactions. This indicates a balance between compaction within the R region, enabling its ability to act as a dynamic integrator and binding to multiple interaction partners, with the need for elements within the R region sequence to be accessible to mediate binding. Further characterization of the phosphorylated R region and comparison to these studies will improve our understanding of these key events at this regulatory centre of CFTR.
Chapter 4: The R Region Interacts with NBD1 Predominantly Via Multiple Transient Helices

The interaction experiments involving labeled R region and unlabeled wild-type NBD1 in this chapter are as published in Jennifer M.R. Baker, Rhea P. Hudson, Voula Kanelis, Wing-Yiu Choy, Patrick H. Thibodeau, Philip J. Thomas, and Julie D. Forman-Kay. CFTR Regulatory (R) Region Interacts with NBD1 Predominantly Via Multiple Transient Helices. (2007) Nature Structural and Molecular Biology, 14, 738-745, a manuscript I wrote. I performed the experiments and analyzed the data presented in this chapter except for the purification of unlabeled wild-type and ΔF508 NBD1 which was performed by Rhea P. Hudson, and the free NBD1 TROSY-HSQC spectrum (Figure 4.4a) and unpublished NBD1 resonance assignments used to analyze the data presented in Figure 4.4b which were courtesy of Voula Kanelis.
4.1 Introduction

Optimal CFTR channel activity requires ATP binding and hydrolysis at the NBDs and PKA phosphorylation at multiple sites, primarily within the R region. Phosphorylation of various sites is generally additive, with the lack of requirement of any single site for channel activity. Previously, there was indirect evidence for an interaction between the R region and NBD1 (Neville, Rozanas et al. 1998; Howell, Borchardt et al. 2004), although the phosphorylation and ATP dependence of the interaction was unknown and there was limited residue-specific data. The ΔF508 mutation alters the CFTR channel response to PKA phosphorylation, suggesting that R region interactions may be modified with ΔF508 NBD1 (Wang, Zeltwanger et al. 2000). To improve our understanding of NBD1 – R region interactions at the molecular level, we performed NMR studies from the perspectives of both the R region and NBD1, including studies of labeled R region with unlabeled wild-type or ΔF508 NBD1 and of labeled wild-type NBD1 with unlabeled R region. Structural and dynamic insights from these studies allow us to propose a dynamic model of R region interactions with NBD1. This model links together regulatory events within cytoplasmic segments of CFTR, producing a signal that can then be transmitted to the channel pore, likely via other CFTR segments including ICL1 (Kanelis, Hudson et al. 2009).

4.2 Experimental procedures

4.2.1 Protein preparation

Human CFTR (P13569) R region was expressed and purified as in section 2.2.1 with either uniform $^{15}$N/$^{13}$C isotopic labeling or unlabeled in BL21(DE3) CodonPlus E. coli grown in M9 media. Selected R region samples were phosphorylated as in section 2.2.2. Both wild-type and ΔF508 murine CFTR (P26361) NBD1 were expressed in the same bacterial strain with either uniform $^{15}$N isotopic labeling in M9 media or unlabeled in LB media from a plasmid encoding residues 389–653 as an N-terminal, His 6x-smt3 (SUMO) fusion protein obtained from our collaborator Philip J. Thomas at the University of Texas Southwestern at Dallas (Lewis, Buchanan et al. 2004). Note that this construct does not contain the “RE” comprising the first ~20 residues of the R region. Murine
NBD1 protein was used since wild-type human CFTR NBD1 is highly insoluble (Lewis, Buchanan et al. 2004). NBD1 was purified from the soluble fraction beginning with Ni\(^{2+}\) affinity chromatography and SUMO protease (Invitrogen) cleavage of the His 6x-SUMO tag. Next, NBD1 was purified using size exclusion chromatography on a HiLoad Superdex 75 column (Pharmacia), followed by further Ni\(^{2+}\) affinity chromatography to remove the His-SUMO tag. Purified R region was added to ∆F508 NBD1 prior to size exclusion chromatography for preparation of the “bound” ∆F508 NBD1 sample (Figure 4.3) to increase ∆F508 NBD1 solubility.

4.2.2 \(^{15}\)N/\(^{13}\)C labeled R region interaction experiments with unlabeled wild-type and ∆F508 mNBD1

HNCO spectra (Sattler, Schleucher et al. 1999) were recorded with 16 transients at 10 °C on 50 µM \(^{15}\)N/\(^{13}\)C R region in the absence and presence of 100 µM (50 µM) unlabeled wild-type (∆F508) mNBD1. Solution conditions were 50 mM NaPO\(_4\) (pH 6.8), 200 mM NaCl, 5 mM ATP, 5 mM MgCl\(_2\), 2% (v/v) glycerol, 2 mM DTT for wild-type ATP-bound NBD1. Solution conditions were the same for wild-type apo-NBD1, with the addition of 2% (w/v) glucose. Solution conditions for ∆F508 ATP-bound NBD1 were the same as wild-type ATP-bound NBD1, except that 4% (v/v) glycerol was used. Buffer conditions were optimized for NBD1 solubility and resulted in small chemical shift changes, compared to spectra of the free \(^{15}\)N/\(^{13}\)C labeled R region from resonance assignment experiments (section 2.2.3). Spectra were processed using NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed by taking the ratio of bound to free peak intensity measured in NMRView (Johnson and Blevins 1994). A uniform scaling factor averaged from several non-interacting residues was applied to the R region peak intensities bound to apo-NBD1 to compensate for a small amount of NBD1 precipitation during the experiment. This precipitation also reduced the amount of R region in solution as evident by the overall decrease in intensity of R region resonances across the full length of the protein.
4.2.3 $^{15}$N labeled wild-type mNBD1 interaction experiments with unlabeled R region

TROSY-HSQC spectra (Pervushin, Riek et al. 1997) were recorded of $^{15}$N labeled wild-type mNBD1 in the absence (86 µM) and presence (300 µM) of 400 µM R region at 20 °C. Different NBD1 concentrations were used because R region binding improved NBD1 solubility and stability, as noted by the addition of free R region to ΔF508 NBD1 preparations (see 4.2.1). Solution conditions were 20 mM NaPO$_4$ (pH 7), 150 mM NaCl, 5 mM ATP, 5 mM MgCl$_2$, 2% (v/v) glycerol, 2 mM DTT, consistent with those used for spectral assignment of a related NBD1 construct (Kanelis, Hudson et al. 2009). Spectra were processed using NMRPipe (Delaglio, Grzesiek et al. 1995) and overlaid in NMRView (Johnson and Blevins 1994).

4.3 Results

4.3.1 NBD1 predominantly binds nonphosphorylated R region

The presence of direct interactions between the isolated R region and NBD1 and the modulation of interactions by phosphorylation and ATP binding, key regulatory events in full length CFTR, were tested by adding unlabeled wild-type NBD1 to $^{15}$N/$^{13}$C-labeled R region samples. The addition of NBD1 primarily leads to reduced peak intensity for specific resonances in the nonphosphorylated R region spectrum, as well as small chemical shift changes (Figure 4.1a). NMR lineshapes reflect the motion of the bond vector measured in the experiment and are sensitive to the overall protein tumbling, as they are dependent on the $R_2$ relaxation rate (described in chapter 3). Disordered proteins have characteristically narrow lineshapes (low $R_2$ relaxation rates), reflecting rapid conformational averaging in the sample and fairly independent motion for each peptide plane. Changes in peak intensity upon interactions reflect contributions from a number of factors, with the largest contribution being direct binding to the target protein, causing the interacting residues of the R region to tumble more slowly due to formation of a complex and leading to broadening of the peak. R region residues not directly at the interaction interface may also appear to be tumbling more slowly due to interactions with other R region residues that directly bind. Changes in interactions within the R region and exchange between free and bound conformations also contribute to the
observed lineshape. Examples of effects on the NMR resonances upon binding NBD1 are shown in Figure 4.1b for Ser790 and Figure 4.1c for Thr760. Residue Ser790 has a bound/free peak intensity ratio greater than 1, reflecting a change in intra-R region contacts upon NBD1 binding, where contacts within the R region present in the free state are released upon NBD1 binding. While it is unknown whether we have completely saturated the binding (i.e. if all of the R region in the sample interacts with NBD1), we refer to the R region samples in the presence of the added NBD1 as "bound".

Figure 4.1 Interaction of the R region with wild-type NBD1

(a) Superposition of a selected portion of $^1$H$^N$-$^{15}$N HSQC spectra of nonphosphorylated $^{15}$N/$^{13}$C-labeled R region in the free state (black) and in the presence of ATP-bound wild-type NBD1 (red), with resolved peaks labeled with their assignments. Note that minor chemical shift changes with NBD1 binding were readily tracked in HNCO spectra. (b) Slice at 115.36 ppm ($^{15}$N frequency) of the HNCO spectrum showing the Ser790 peak, the only one that maintains intensity in this experiment upon addition of NBD1. (c) Slice at 115.07 ppm ($^{15}$N frequency) of the HNCO spectrum showing the Thr760 peak, which is weaker upon NBD1 binding, indicative of interactions with NBD1. In (b) and (c), the peak in the free R region spectrum is shown (left, in black) and in the spectrum of the R region with the addition of NBD1 (right, in red). A trace through the approximate center of the peak (dashed line representing a slice through the 3D data) is shown at the bottom, illustrating the lineshape.

The ratios of the peak intensities of resonances of the NBD1 "bound" R region to those of the free R region are shown in Figure 4.2 for nonphosphorylated and phosphorylated
R region and wild-type NBD1 in the absence and presence of ATP. If no interactions were observed, the ratio would be one for all residues. A relatively uniform and low ratio, with only minor deviations, would be expected for all residues if the R region became ordered upon binding, as the entire R region would largely tumble slowly as a unit in complex with NBD1 rather than nearly independently. If very short stretches of extended structure were involved in the interaction, the ratios would dip with sharp minima. Instead, what is observed are ratios that vary from 0 to 1.3, with broad rather than sharp minima, reflective of longer stretches of residues interacting with varying affinities and no global disorder-to-order transition. Multiple R region segments appear to bind NBD1, to varying degrees, implying dynamic exchange of multiple R region binding segments on and off NBD1.
**Figure 4.2** Analysis of R region interactions with wild-type NBD1

The ratios of $^{15}$N/$^{13}$C-labeled R region peak intensities with and without NBD1 from HNCO experiments are plotted. Nonphosphorylated (a) and phosphorylated (b) R region interactions with apo-NBD1. Nonphosphorylated (c) and phosphorylated (d) R region interactions with ATP-bound NBD1. Phosphorylation sites are shown with red circles. Bars are colored in each panel (a)−(d) to reflect secondary structural properties of the R region in the absence of NBD1 (Figure 2.5) where at least three consecutive residues have SSP values above 5% (indicating fractional α-helical structure, yellow) or SSP values below −5% (indicating fractional β-strand structure, blue), and the remaining residues are colored grey. Error bars are derived from error propagation of the noise level of the spectrum. Bound peak intensities were calculated at the chemical shifts of the bound peak, as some peaks had small chemical shift changes upon interaction with NBD1. Note that the absence of bars indicates that data could not be analyzed for those residues, with the exception of zero values of the bound to free peak intensity for residues 765, 771, 774 (a) and residues 763, 766, 771, 772, and 775 (c).

In the case of the nonphosphorylated R region, multiple segments of the sequence are involved in interactions with apo- and ATP-bound NBD1, including the major interaction site surrounding Ser768 for which some resonances are broadened to zero intensity (Figure 4.2a,c). The majority of these interaction sites either include, or are just C-terminal to, the demonstrated in vivo phosphorylation sites at 660, 700, 737, 795, and 813 (Cohn, Nairn et al. 1992; Picciotto, Cohn et al. 1992; Chang, Tabcharani et al. 1993). The interaction of residues 661-671 with NBD1 is consistent with the crystal structures of NBD1 containing these residues as part of the regulatory extension (RE) contacting the NBD1 core (Lewis, Buchanan et al. 2004; Lewis, Zhao et al. 2005; Thibodeau, Brautigam et al. 2005). In the context of full length CFTR, in which NBD1 and the R region are covalently linked, the affinities of the individual interacting segments would be expected to be even higher than those shown here with NBD1 and the R region as separate polypeptides. Phosphorylation of the R region, however, abrogates most binding to NBD1 (Figure 4.2b,d), with the exception of weaker binding in the vicinity of Ser768 when ATP is present (Figure 4.2d). Although ATP binding by NBD1 causes few changes in the interaction of NBD1 and the nonphosphorylated R region, ATP binding enhances the interaction of the Ser768 site when the R region is phosphorylated. In the presence of ATP, the maintenance or increase in peak intensity near Ser790 compared to other residues is observed for both phosphorylation states
(Figure 4.2c,d), indicative of interactions present within the free state of the R region involving Ser790 that are released upon interaction with NBD1.

4.3.2 Nonphosphorylated R region binds ΔF508 NBD1 similarly to wild-type NBD1

Differences in R region binding to wild-type versus ΔF508 NBD1 are suggested by electrophysiological evidence (Wang, Zeltwanger et al. 2000) and structural studies using NBD1-RE fragments that include the first 30-40 residues of the R region (Kanelis, Hudson et al. 2009). Differences were examined at the molecular level for nonphosphorylated R region binding to ATP-bound ΔF508 NBD1, however limited ΔF508 NBD1 solubility did not permit testing the phosphorylation and ATP-binding dependence of this interaction. The addition of unlabeled ATP-bound ΔF508 NBD1 to 15N/13C labeled R region leads primarily to reduced peak intensity as well as to minor chemical shift changes as monitored in HNCO spectra, similar to the addition of wild-type ATP-bound NBD1. These changes were analyzed by examining the ratio of bound to free R region peak intensity (Figure 4.3a). This plot shows broad minima, where stretches of nonphosphorylated R region residues bind to ΔF508 NBD1 to varying degrees, implying a dynamic R region interaction with ΔF508 NBD1 where R region binding segments bind and are released from NBD1. The residues of the nonphosphorylated R region involved in binding ΔF508 and wild-type NBD1 are mostly similar (compare Figure 4.3a and Figure 4.3b), including near residues 660-670, 700-715, 730-740, and 750-778. There are differences near residues 803-816 and 828-836, which appear to bind wild-type more than ΔF508 NBD1. In addition, the high bound-to-free peak intensity ratios seen for K786 and S790 with wild-type NBD1 binding are not seen with ΔF508 NBD1 binding, indicating that intra-R region contacts for these residues are not affected in the same manner. The two studies have different molar ratios of R region to NBD1, as limited solubility of ΔF508 NBD1 permitted only a 1:1 molar ratio, while the wild-type NBD1 experiments were recorded at a 1:2 molar ratio. In both cases, experiments were recorded in the absence and presence of NBD1, but without recording multiple NBD1 titration points because of limited NBD1 stability. As a result, the proportion of free to bound R region is unknown within and between experiments and it is difficult to quantitatively compare the two experiments as their
relative affinities and the proportion of bound R region are unknown. However, it does appear that the pattern of R region binding sites is overlapping but not identical between wild-type and ΔF508 NBD1.

**Figure 4.3** Comparison of nonphosphorylated R region interactions with wild-type and ΔF508 NBD1

The ratios of $^{15}$N/$^{13}$C-labeled nonphosphorylated R region peak intensities with and without ATP-bound -
(a) ΔF508 and (b) wild-type NBD1 (same as Figure 4.2c) from HNCO experiments are plotted. Phosphorylation sites are shown with red circles. Error bars are derived from error propagation of the noise level of the spectrum. Bound peak intensities were calculated at the chemical shifts of the bound peak, as some peaks had small chemical shift changes upon interaction with NBD1. Note that the absence of bars indicates that data could not be analyzed for those residues with the exception of zero values of the bound to free peak intensity for residues 763, 766, 771, 772, and 775 (b).

### 4.3.3 NBD1 spectral changes upon nonphosphorylated R region binding

Having examined interactions between the R region and NBD1 as determined from R region spectra, we also wanted to probe these interactions from the perspective of
NBD1. We added unlabeled nonphosphorylated R region to $^{15}\text{N}$-labeled NBD1 samples and monitored spectral changes. In the free state, NBD1 resonances have varying intensities. NBD1 is folded, as the overall spectral dispersion in the $^1\text{H}$ dimension is quite disperse, however there are many sharp resonances in the centre of the spectrum near $\sim$ 8.2 ppm from flexible segments in NBD1, likely including the regulatory insert (RI) that samples disordered conformations and only transiently binds the NBD1 core (Figure 4.5a) (Lewis, Buchanan et al. 2004; Lewis, Zhao et al. 2005; Thibodeau, Brautigam et al. 2005; Kanelis, Hudson et al. 2009). Numerous other resonances at chemical shifts corresponding to residues in the folded core are broadened, indicating the presence of $\mu$s-ms timescale motions within the NBD1 core. With the addition of R region, many NBD1 resonances exhibit spectral changes. Because of the limited solubility of NBD1, these experiments examine NBD1 in the presence of a 1.25 molar excess of R region, without titrating the R region into NBD1 or being able to demonstrate saturation of binding. As a result, the ‘bound’ NBD1 spectrum shown in Figure 4.4a likely represents an intermediate state with both free and bound NBD1 present.
Figure 4.4 Nonphosphorylated R region binding to NBD1

(a) Superposition of TROSY-HSQC spectra of $^{15}$N-labeled ATP-bound NBD1 in the free state (86µm NBD1 in black, blue peaks) and in the presence of unlabeled, nonphosphorylated R region (300µm NBD1 and 400µm R region in red, magenta peaks). Many chemical shift and lineshape changes occur with the addition of R region (selected peaks indicated with arrows), indicating R region binding. (b) Spectral changes are mapped onto the NBD1-RE ATP-bound structure (1R0X.pdb) displayed without the RE (Lewis, Buchanan et al. 2004), based on spectral assignments of a related NBD1 construct (Kanelis, Hudson et al. 2009). Residues with spectral changes are indicated in magenta, those without are indicated in aqua. Gray/white indicates residues that could not be assigned. The F508 residue also has spectral changes and is indicated in gold. (left) NBD1 structure in the same orientation as in Figure 1.10, (right) rotated by 180° about the indicated axis. The NBD1 surface is rendered with MOLMOL (Koradi, Billeter et al. 1996).

Spectral assignments of a related construct (Kanelis, Hudson et al. 2009) allow assignment of ~30% of the NBD1 spectrum. Mapping these spectral changes to the 1R0X.pdb NBD1 crystal structure (Lewis, Buchanan et al. 2004; Lewis, Zhao et al. 2005; Thibodeau, Brautigam et al. 2005) indicates that more than one NBD1 surface is affected by R region binding (Figure 4.4b). This may result from the R region directly binding at more than one NBD1 surface(s) or from NBD1 conformational changes upon R region binding. Limited spectral assignments (Kanelis, Hudson et al. 2009), difficulty in transferring assignments from the NBD1 construct used for spectral assignment to the one used in the current study, and uncertainty in the degree of saturation of binding preclude more specific characterization.

4.4 Discussion

The R region regulates CFTR in a complex manner, dependent on phosphorylation of multiple PKA sites and sensitive to a variety of interaction partners. To aid in understanding the molecular basis for this mode of regulation, we have characterized interactions of the R region with NBD1, site of the ΔF508 mutation and ATP binding/hydrolysis. Interpreting these results requires particular attention to the unique properties of disordered protein segments necessary to explain R region behaviour.

The function of intrinsically disordered protein regions in protein recognition has been suggested to be via two predominant mechanisms. There are many examples of disordered proteins that undergo a disorder-to-order transition upon binding, stabilizing
a structural domain within a previously disordered region (for a review see (Dyson and Wright 2002)). Alternatively, primary sequence motifs (approximately 5–10 residues) that are targets for modular binding domains such as SH2, SH3, and WW domains are present largely within disordered regions of proteins and their interactions involve an extended chain (or turn) binding to the interface of the domain (Pawson 1995). The presence of fractional helical structure in the R region free state and changes in resonance intensity over multiple stretches of 10–15 residues upon binding to NBD1 provide evidence that R region interactions are mediated by stabilization of fluctuating helical structural elements, without a complete disorder-to-order transition upon binding, representing an intermediate between these two mechanisms (Fuxreiter, Simon et al. 2004; Oldfield, Cheng et al. 2005). If a complete disorder-to-order transition occurred, all the resonances of an R region with substantial tertiary contacts would be affected to a similar degree upon binding to the larger NBD1, and the values in Figure 4.2a–d would be similar. Instead, multiple segments of the R region sequence show different effects upon addition of NBD1, indicating that multiple R region segments bind and dissociate from NBD1 in a dynamic fashion, with the intensities observed reporting the population-weighted average degree of association and relative affinity of each segment. Binding of residues 763–777 is tightest in the nonphosphorylated R region, with almost complete disappearance of the resonances upon binding to NBD1. These residues partially overlap with the segment 771–779, comprising a pattern indicative of an α-helical molecular recognition element (α-MoRE) (Oldfield, Cheng et al. 2005), elements that are disordered in the free state and undergo helix stabilization in the bound state.

The coincidence of fractional helical structure in the free R region (indicated by yellow bars) and NBD1 interacting regions is highlighted in Figure 4.2. Transient helices appear to be stabilized upon NBD1 binding, extending the lengths of the helices and their populations, as evidenced by the broad minima observed. The inexact correspondence between free state helical structure and binding reflects this structural adjustment upon binding as well as electrostatic and other energetic contributions to the interaction. The correlation of reduction in helical structure for many R region segments (Figure 2.5) and loss of most NBD1 binding (Figure 4.2) upon phosphorylation is also
indicative of the role of helices in the interaction, with additional support coming from the helical RE interacting with NBD1 in crystal structures of NBD1-RE (Lewis, Buchanan et al. 2004), and the presence of an α-MoRE sequence and the highest $R_2$ relaxation rates (Figure 3.1a-b) for residues 763-777, the most stabilized helix (Oldfield, Cheng et al. 2005). Together, these data provide evidence for a dynamic complex in which multiple fluctuating helices present in the free R region bind NBD1, are transiently stabilized at the interaction interface, and are subsequently released.

Multiple R region segments are involved in ΔF508 NBD1 binding, substantially overlapping with those involved in wild-type NBD1 binding. However, the ΔF508 mutation removes some R region – NBD1 interactions near the R region C-terminus. These changes in nonphosphorylated R region interactions may contribute to the altered ΔF508 CFTR response to PKA phosphorylation noted in electrophysiological experiments (Wang, Zeltwanger et al. 2000), but effects from altered phosphorylated R region interactions with NBD1 are likely. Comparing NMR spectra of wild-type and ΔF508 NBD1-RE, PKA phosphorylated at residues 422 in the RI and 660 and 670 in the RE, shows altered dynamics with the ΔF508 mutation (Kanelis, Hudson et al. 2009). This is consistent with altered binding and release of the RI and RE, behaviour that likely extends to the rest of the R region.

This model in which multiple R region segments can bind NBD1 in the context of isolated constructs may explain the lack of requirement for specific phosphorylation sites in the regulation of full length CFTR (Cheng, Rich et al. 1991; Chang, Tabcharani et al. 1993). The PKA phosphorylation sites are contained within segments of the R region that interact with NBD1 (Figure 4.2), so mutation of a subset of phosphorylation sites in the context of full length CFTR would likely primarily effect CFTR regulation by their coincident interacting segments. The remaining wild-type phosphorylation sites and their surrounding interaction segments could still act, allowing channel regulation by PKA consistent with partial channel activity. Milder phenotypes are seen for many CF-causing CFTR missense mutations within the R region, consistent with this multisite behavior, with the majority of these mutations found within the PKA recognition and phosphorylation sites (R709N, S712C, R735K, S737F, V754M, R766M, R810G, S813P) (http://www.genet.sickkids.on.ca/cftr/app).
Two possible models describe R region binding to NBD1 consistent with the current data (Figures 4.2, 4.4). A simple model of discrete R region segments binding to discrete NBD1 surface(s) is inconsistent with the observed differential binding of multiple R region segments to NBD1 (Figure 4.2). The many spectral changes in $^{15}$N labeled NBD1 upon the addition of unlabeled R region result from direct R region binding and conformational/dynamic changes within NBD1 (Figure 4.4). One possible model (Figure 4.5a) consists of multiple R region segments dynamically interacting with varying affinities with a single NBD1 binding surface, accompanied by larger changes in NBD1 conformation/dynamics as indicated by the spectral changes that map to residues throughout the NBD1 structure (Figure 4.4b). A second possibility (Figure 4.5b) includes multiple R region segments binding to multiple NBD1 surfaces in a highly dynamic equilibrium. NBD1 resonances corresponding to residues at these interfaces are affected, as well as others throughout the structure reflecting potentially more limited structural/dynamic changes propagated from the interaction interfaces.

**Figure 4.5 Possible modes of NBD1-R region interactions**

Possible interaction models consistent with NBD1 spectral changes upon R region binding, indicating the R region (red curve), R region binding segments (orange spheres), NBD1 (multi-colour crescent), and NBD1 binding surfaces (grey ellipses). Black arrows indicate the dynamic exchange of R region binding segments with NBD1 surface(s). (a) Multiple R region segments bind to a single NBD1 surface in a highly dynamic fashion. The blue to orange gradient indicates the presence of large changes in NBD1 conformation/dynamics. (b) Multiple R region segments bind to multiple NBD1 surfaces in a highly dynamic equilibrium. Potentially more limited changes in NBD1 dynamics are indicated by a blue to green gradient.

A possible NBD1 surface for interaction with the R region was revealed in NBD1-RE crystal structures (Lewis, Buchanan et al. 2004; Lewis, Zhao et al. 2005; Thibodeau,
Brautigam et al. 2005). The residues on this surface could not be assigned due to broadening reflecting motion of the RE (Kanelis, Hudson et al. 2009). The RE (residues 655-673) forms a helix that packs against NBD1 along the putative NBD1-NBD2 dimerization interface in many of the structures (Lewis, Buchanan et al. 2004; Lewis, Zhao et al. 2005; Thibodeau, Brautigam et al. 2005), demonstrating how the R region could inhibit CFTR function by blocking NBD1/NBD2 association, nucleotide hydrolysis, and channel opening. An inhibition of NBD1-NBD2 dimerization is also supported by the findings that phosphorylation increases dimerization of NBD1-R-GST and NBD2-MBP (Howell, Borchardt et al. 2004) \textit{in vitro} and increases NBD1-NBD2 crosslinking (Mense, Vergani et al. 2006) in the context of full length CFTR. Based on our data, the association of the RE helix with NBD1 along the dimerization interface is transient, consistent with another NBD1-RE structure with a different crystallized conformation from this structural ensemble in which the RE is rotated away from the dimerization interface (Figure 1.10) (Lewis, Zhao et al. 2005). Studies of CFTR composite channels missing the RE showed largely wild-type function (Csanady, Chan et al. 2005), supporting a model of transient association of multiple R region segments to NBD1 in full length CFTR. The RE helix observed in crystal structures determined in various nucleotide-bound states (Lewis, Buchanan et al. 2004) is an amphipathic helix, packing against a primarily hydrophobic surface of NBD1 and burying 880Å² of surface area. Based on helical wheel analysis, residues 733–739 and 761–777 (depending on the side chain conformation of Arg766) are also candidates to form amphipathic helices, which may bind the same or other NBD1 surface(s) as the RE in a dynamic equilibrium. Additional R region residues, including 767–780 that interact with ATP-bound NBD1 independently of phosphorylation, may bind to this same NBD1 surface or at other NBD1 surface(s) (Ma, Zhao et al. 1997).
Figure 4.6 Schematic illustrating phosphorylation-induced structural changes in the R region and consequent redistribution of binding equilibria with various intramolecular regulatory interaction partners.

R region is shown as a red curve, with multiple helices reflecting the fractional helical structure in portions of the sequence. Gray ellipses, putative binding surfaces of interaction partners; green arrows, potential tertiary interactions within the R region; gold arrows, interactions with binding partners (multiple arrows represent dynamic exchange of both multiple R region binding sites and multiple interactions interfaces, without implying involvement of specific R region segments). Other components of CFTR are also shown, including membrane-spanning domains (MSDs), intracellular domains (ICDs), nucleotide binding domains (NBDs), and cytoplasmic, helical N terminus (N-tail). (a,b) Interactions favoured by nonphosphorylated (a) and phosphorylated (b) R region. Upon phosphorylation R region has less helical structure and reduced interactions with NBD1 and possibly NBD2, as well as within the R region; NBD1-NBD2 dimerization leads to channel opening.

Both "switch" and "rheostat" mechanisms are known modes of regulation via phosphorylation of disordered proteins. Multiple phosphorylated sites of the intrinsically disordered cyclin-dependent kinase inhibitor Sic1 bind to a single site on the Cdc4 component of an SCF ubiquitin ligase in a dynamic equilibrium to allow high affinity
“switch-like” binding (Nash, Tang et al. 2001; Mittag, Orlicky et al. 2008). In contrast, the intrinsically disordered N-terminal segment of the Ets-1 transcription factor has multiple phosphorylation sites that act as a rheostat to control autoinhibition of the ETS domain (Pufall, Lee et al. 2005). The R region may be a similar rheostat, however in this case primarily with the nonphosphorylated state binding. Regulation of CFTR function occurs by phosphorylation-dependent modulation of structural properties at a large number of potential interacting segments, providing a mechanism for a graded CFTR response to PKA binding and/or phosphorylation (Chang, Tabcharani et al. 1993; Wilkinson, Strong et al. 1997; Vais, Zhang et al. 2004). Moderate channel activity is possible with low levels of phosphorylation, however additional phosphorylation further increases CFTR activity (Chang, Tabcharani et al. 1993). Phosphorylation of various R region sites is generally additive, with the lack of requirement of any single site for channel activity (Cheng, Rich et al. 1991; Chang, Tabcharani et al. 1993) and variability in the effect on channel activity for mutations at different phosphorylation sites (Wilkinson, Strong et al. 1997; Vais, Zhang et al. 2004). Transient R region interaction segments with different affinities for NBD1 may act together to shift the equilibrium between NBD1 monomer and NBD1-NBD2 heterodimer, as well as with other binding partners (Figure 4.6). In this model, the nonphosphorylated R region inhibits NBD dimerization by binding to NBD1. Dynamic interconversion between various R region segments binding and dissociating from NBD1 allows PKA to access the R region PKA consensus motifs, enabling phosphorylation and leading to an equilibrium shift away from binding NBD1. In addition to facilitating NBD1/NBD2 dimerization, this enhances R region binding to the N-terminus, acting to lengthen the channel open duration. Further NMR experiments to probe the specific residues of the R region involved in interactions with the N-terminus, the NBD2, and other potential interacting segments within CFTR will shed more light on the detailed mechanisms of this disordered phospho-regulatory R region in controlling CFTR function.
Chapter 5: Characterizing the SLC26A3 STAS Domain and Interactions with the R region

The comparison of free wild-type and ΔY526/7 STAS domain in this chapter is as published in Michael R. Dorwart, Nicholi Shcheynikov, Jennifer M.R. Baker, Julie D. Forman-Kay, Schmullem Muallem, and Philip J. Thomas. Congenital Chloride-losing Diarrhea Causing Mutations in the STAS Domain Result in Misfolding and Mistrafficking of SLC26A3. (2008) Journal of Biological Chemistry, 283, 8711-8722. I wrote the sections of this paper pertaining to the data presented here. I performed and analyzed all experiments presented in this chapter.
5.1 Introduction

In addition to regulation of CFTR channel activity by ATP binding/hydrolysis and phosphorylation, additional modulation of channel activity also occurs through interaction with other proteins, including SLC26A3 (Ko, Zeng et al. 2004). The SLC26A (SoLute Carrier family 26) is a recently described membrane protein family containing 10 human proteins (Lohi, Kujala et al. 2000; Mount and Romero 2004), including SLC26A3. This protein is predicted to contain between 10 and 14 transmembrane-spanning α-helices and a C-terminal domain (Hoglund, Haila et al. 1996; Kere, Lohi et al. 1999; Moseley, Hoglund et al. 1999; Makela, Kere et al. 2002) that is highly conserved and important for proper ion transport activity (Chernova, Jiang et al. 2003; Shibagaki and Grossman 2004; Shibagaki and Grossman 2006). This C-terminal, Sulfate Transporter and Anti-Sigma (STAS) domain was originally identified by its similarity to bacterial anti-sigma-factor antagonists (Aravind and Koonin 2000), with an additional InterVening Sequence (IVS) or large insert in the primary sequence.

Mutations found in the SLC26A3 gene have been linked to congenital chloride-losing diarrhea (CLD) (Hoglund, Haila et al. 1996), a disease in which patients suffer from watery diarrhea containing elevated Cl\(^-\) concentrations that can prove fatal if left untreated (Holmberg 1986), demonstrating the critical role in fluid regulation of the SLC26A family members. Interestingly, SLC26A3 has been shown to activate CFTR by increasing its open probability six-fold, and, at the same time, SLC26A3 is activated by CFTR, via a “reciprocal” regulatory interaction (Ko, Zeng et al. 2004). This dual activation occurs via the association of the C-termini of both CFTR and SLC26A3 with a PDZ domain-containing protein as well as the binding of the R region to the STAS domain (Wang, Yue et al. 2000; Lamprecht, Heil et al. 2002; Ko, Zeng et al. 2004) (Figure 1.12). The interaction also occurs in the absence of the PDZ-containing protein when CFTR and SLC26A3 are over-expressed. R region phosphorylation enhances the interaction of CFTR with SLC26A3 and results in an increase in activity for both proteins (Ko, Zeng et al. 2004). This reciprocal regulation provides a model for understanding the role of CFTR in the physiologically observed low bicarbonate levels of pancreatic fluid in CF patients.
To gain insight into the molecular mechanisms of both CLD and SLC26A3 regulation of CFTR, structural studies were carried out at residue-specific resolution using NMR. While over 30 mutations in SLC26A3 have been linked to CLD (Makela, Kere et al. 2002), 4 of these mutations reside in the STAS domain. Dorwart and coworkers examined these 4 mutations in the context of full length SLC26A3 and isolated STAS domain. Two mutations, I675/6ins and G702Tins, likely disrupt the STAS domain directly, while two other mutations, ΔY526/7 and I544N, likely disrupt interdomain contacts (Dorwart, Shcheynikov et al. 2008). The ΔY526/7 mutant was chosen for more extensive structural study using NMR to confirm that the primary defect of this mutation is not disruption of the STAS domain. The interaction between phosphorylated R region and wild-type STAS domain was also studied to determine the R region residues involved, providing insight into the possible competition between various partners for binding specific R region segments.

5.2 Experimental procedures

5.2.1 Protein expression and purification

Both wild-type and ΔY526/7 human SLC26A3 (NP000102) STAS domains were expressed in BL21(DE3) CodonPlus RIL (Stratagene) cells with either uniform $^{15}$N isotopic labeling in M9 media supplemented with 5% $^{15}$N-labeled BioExpress media (Cambridge Isotope Laboratories) or unlabeled in LB media from a plasmid encoding residues 510-741 as an N-terminal, His 6x-smt3 (SUMO) fusion protein. Cells were grown at 37 °C to an $A_{600}$ of 0.8, then induced with 750 µM isopropyl β-D-1-thiogalactopyranoside (BioShop) and shifted to 15 °C for overnight expression. STAS domain was purified from the soluble fraction beginning with Ni$^{2+}$ affinity chromatography and SUMO protease (Invitrogen) cleavage of the His 6x-SUMO tag. Ammonium sulfate (750 mM) was added to the cleaved protein solution, then STAS domain was purified using butyl hydrophobic interaction chromatography (Amersham Biosciences). Size exclusion chromatography on a HiLoad Superdex 200 column (Pharmacia) was used for final purification. Human CFTR (P13569) R region was expressed and purified as in section 2.2.1 with uniform $^{15}$N/$^{13}$C isotopic labeling in BL21(DE3) CodonPlus E. coli grown in M9 media. R region samples were
phosphorylated as in section 2.2.2; phosphorylated serine residues were verified by their chemical shifts in NMR spectra.

5.2.2 Comparison of $^{15}$N-labeled Wild-type and ΔY526/7 STAS domains

TROSY-HSQC spectra (Pervushin, Riek et al. 1997) of 90 µM wild type and ΔY526/7 STAS domains were collected using a 600-MHz Varian Inova spectrometer with a pulsed-field gradient, triple resonance cold probe at 30 °C for 12 h each. Solution conditions were 50 mM potassium phosphate (pH 6.8), 400 mM potassium chloride, 100 mM arginine, 2% glucose, 2 mM DTT. Spectra were processed using NMRPipe (Delaglio, Grzesiek et al. 1995) and displayed in NMRView (Johnson and Blevins 1994).

5.2.3 $^{15}$N/$^{13}$C labeled R region interaction experiments with unlabeled wild-type STAS domain

HNCO spectra (Sattler, Schleucher et al. 1999) were recorded with 16 transients at 10 °C on 50 µM $^{15}$N/$^{13}$C phosphorylated R region in the absence and presence of 150 µM unlabeled wild-type STAS domain. Solution conditions were 50 mM K$^{+}$ phosphate (pH 6.8), 400 mM KCl, 100 mM Arg, 2% glucose, 2 mM DTT. Buffer conditions were optimized for STAS solubility and resulted in small chemical shift changes, compared to spectra of the free $^{15}$N/$^{13}$C labeled R region from resonance assignment experiments (section 2.2.3). Spectra were processed using NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed by taking the ratio of bound to free peak intensity measured in NMRView (Johnson and Blevins 1994).

5.3 Results

5.3.1 ΔY526/7 mutation does not change STAS domain structure

The molecular consequences of the ΔY526/7 mutation, known to cause CLD, on SLC26A3 structure and function are not known. Electrophysiology and cell biology experiments performed by Dorwart and coworkers (Dorwart, Shcheynikov et al. 2008) in the laboratory of Philip Thomas showed that full length SLC26A3 with the ΔY526/7 mutation has decreased Cl$^{-}$/HCO$_3$$^{-}$ exchange and lower levels of complex glycosylation in comparison to wild-type, indicating that this mutant is inefficiently trafficked to the cell membrane. To determine if this effect is primarily because of STAS domain misfolding
or because of altered inter-domain contacts, we probed changes in domain structure by comparing HSQC spectra of wild-type and ΔY526/7 STAS domains (Figure 5.1).

Figure 5.1 Comparison of wild-type and ΔY526/7 STAS domain TROSY-HSQC spectra

$^1$H$^N$, $^{15}$N TROSY-HSQC NMR spectrum of the human SLC26A3 STAS domain, comprising residues 510-741 with wild-type (black) and ΔY526/7 (red) superimposed.

The TROSY-HSQC spectra of wild-type and ΔY526/7 STAS domains show nitrogen-proton cross-peaks with chemical shift dispersions typical of a well folded protein, with peaks in the center of the spectrum likely from flexible loops, termini, or the IVS. The two spectra overlay well with only minor differences, consistent with absence of peaks and local chemical shift changes near the site of the mutation. The wild-type and ΔY526/7 STAS domains are structurally similar; indicating that the primary effect of this mutation is likely altered inter-domain contacts within SLC26A3.
5.3.2 Phosphorylated R region interaction with STAS domain

The phosphorylated R region has been shown to interact with the STAS domain of SLC26A3, with a resulting increase in CFTR and SLC26A3 channel activity (Ko, Zeng et al. 2004). To begin characterizing this interaction at the molecular level, spectral changes in phosphorylated $^{15}$N/$^{13}$C R region were monitored with the addition of unlabeled STAS domain. The addition of STAS domain primarily led to changes in R region peak intensity, quantified by taking the ratio of bound to free peak intensity (Figure 5.2). The overall changes in peak intensity were smaller than for the phosphorylated R region interaction with NBD1 (Figure 4.2), even with a three-fold molar excess of STAS, likely because interactions were screened by the high salt buffer optimized for STAS solubility. Nonetheless, similar phosphorylated R region segments are involved in STAS domain and NBD1 binding, mostly involving the segment near S768, as well as minor binding near residue 805. Both of these R region segments maintain helical propensity in the phosphorylated state (although with a reduced propensity compared to nonphosphorylated R region near 805) (Figure 2.5), indicating that transiently helical segments may mediate interactions. The similar segments involved in binding STAS and NBD1 likely indicate a direct competition for the two in binding phosphorylated R region. Binding of nonphosphorylated R region to STAS should also be probed in future experiments to determine if residual STAS binding may also compete with NBD1 in the nonphosphorylated state.
Figure 5.2 Analysis of phosphorylated R region interactions with STAS and NBD1

The ratios of 50 µM $^{15}$N/$^{13}$C-labeled phosphorylated R region peak intensities with and without (a) 150 µM STAS domain or (b) 100 µM ATP-bound NBD1 (data as in Figure 5.4d) from HNCO experiments are plotted. Bars are colored in each panel (a)−(b) to reflect secondary structural properties of the R region in the absence of STAS or NBD1 (Figure 2.5) where at least three consecutive residues have SSP values above 5% (indicating fractional α-helical structure, yellow) or SSP values below −5% (indicating fractional β-strand structure, blue), and the remaining residues are colored grey. Phosphorylation sites are shown with red circles. Error bars are derived from error propagation of the noise level of the spectrum. Note that the absence of bars indicates that data could not be analyzed for those residues.

5.4 Discussion

SLC26A3 is a key protein in maintaining the proper physiological balance of ions in the gut, and may also play an important role in maintaining alkaline conditions in the pancreas (Hoglund, Haila et al. 1996; Ko, Zeng et al. 2004). Mutations in SLC26A3 cause CLD, with several CLD-causing mutations residing in the STAS domain including the ΔY526/7 mutation. This mutation does not grossly affect the structure of the STAS
domain (Figure 5.1), suggesting a role for this STAS domain surface in key inter-domain interactions important for SLC26A3 protein folding and channel activity (Dorwart, Shcheynikov et al. 2008).

**Figure 5.3** Schematic illustrating phosphorylation-induced structural changes in the R region and consequent redistribution of binding equilibria with various regulatory interaction partners. 

R region is shown as a red curve, with multiple helices reflecting the fractional helical structure in portions of the sequence. Gray ellipses, putative binding surfaces of interaction partners; green arrows, potential tertiary interactions within the R region; gold arrows, interactions with binding partners (multiple arrows represent dynamic exchange of both multiple R region binding sites and multiple interactions interfaces, without implying involvement of specific R region segments). Other components of CFTR are also shown, including membrane-spanning domains (MSDs), intracellular domains (ICDs), nucleotide binding domains (NBDs), and cytoplasmic, helical N terminus (N-tail). The SLC26A3 chloride/bicarbonate transporter is shown with its STAS domain. PKA and protein phosphatase 2A (PP 2A) are also shown. (a,b) Interactions favoured by nonphosphorylated (a) and phosphorylated (b) R region. Upon phosphorylation R region has less helical structure and reduced interactions with NBD1 and possibly NBD2, as well as within the R region, while it has increased interactions with the N-tail, PP 2A and STAS domain; NBD1-NBD2 dimerization leads to channel opening.

The SLC26A3 STAS domain also interacts with the phosphorylated R region, increasing the activity of both SLC26A3 and CFTR (Ko, Zeng et al. 2004). The R region segment near S768 interacts most strongly with both STAS and NBD1, indicating that the two
proteins may compete for phosphorylated R region binding. The R region segment near S768 is functionally important in full length CFTR, as removing residues 760-783 causes the channel to open without PKA phosphorylation (Baldursson, Ostedgaard et al. 2001). This R region segment may bind to a different NBD1 surface than NBD1/NBD2 dimerization interface and involve regulatory events in addition to the R region blocking dimerization. Preferential binding of this segment to STAS in a phosphorylation dependent manner could increase CFTR channel activity by preventing the R region from binding to this other NBD1 surface. Possible direct competition between different interaction partners for binding the same R region segments in the same phosphorylation state expands our previous model (Figure 4.6) to one where R region segments come on and off of their targets and exchange binding partners, including with other proteins (Figure 5.3) like the STAS domain (Ko, Zeng et al. 2004). PKA (Picciotto, Cohn et al. 1992) and protein phosphatase 2A (PP2A) (Vastiau, Cao et al. 2005) could also be involved in this exchange, with the additional effect that these interactions lead to phosphorylation or dephosphorylation, hence moving the overall interaction equilibrium. In the context of full-length CFTR and its physiological binding partners, the large, approximately 200-residue R region and its binding sites may be interacting with more than one binding partner at any one time, acting as a true integrator of stimuli from multiple sources. There are certain to be other inter- and intramolecular interaction partners of the R region still to be identified that control this equilibrium shift. Further NMR experiments to probe the specific residues of the R region involved in these interactions will expand our understanding of this dynamic equilibrium modulating CFTR function.
Chapter 6: Summary and Future Directions
6.1 Summary

This work has examined properties of the nonphosphorylated and phosphorylated R region in the free state, as well as interactions with NBD1 and the SLC26A3 STAS domain. The intrinsically disordered nonphosphorylated and phosphorylated R region of CFTR have been characterized at residue-level resolution, primarily using NMR. HSQC experiments show limited chemical shift dispersion for both nonphosphorylated and phosphorylated R region in the proton dimension, indicating that the R region is intrinsically disordered in solution and that no global folding event occurs upon phosphorylation. The backbone resonances and sidechain carbon nuclei were assigned for nonphosphorylated and phosphorylated R region using NMR triple resonance experiments. SSP calculations show that the nonphosphorylated R region has multiple segments with helical propensity, with values up to approximately 30% populated helix for some residues, with these largely corresponding to segments near the PKA phosphorylation sites. Helical propensity is reduced upon phosphorylation, with some segments also showing increased β–strand propensity, indicating that phosphorylation acts as a structural switch.

To further characterize the free nonphosphorylated and phosphorylated R region, a variety of structural probes were examined. Motions on the fast timescale were probed using $R_1$, $R_2$ and heteronuclear NOE NMR relaxation experiments, indicating the presence of structural contacts in many R region segments that also have higher SSP values. The hydrodynamic radii, probed using NMR PFG experiments, indicate that the R region has intermediate properties to those expected for a fully folded protein and for a completely disordered protein, with the phosphorylated R region being slightly more compact. The nonphosphorylated R region was further characterized by using SAXS to probe the molecular dimensions and measuring RDCs to determine the orientation of N-H bond vectors with respect to a common reference. Inter-residue distances probed using PRE experiments with spin-label probes at six different sites in the primary sequence indicate the presence of transient contacts throughout the R region primary sequence. The program ENSEMBLE (Marsh, Neale et al. 2007) was used to calculate a representative pool of conformations using information from all of the experimentally
measured structural probes. The nonphosphorylated R region ensemble of conformers shows clusters of R region contacts in several segments, including near the N- and C-termini, as well as near S790, a residue that may have contacts with other R region residues that are released upon NBD1 binding (Baker, Hudson et al. 2007). Transiently helical R region segments are relatively uninvolved with R region binding, potentially allowing them to be accessible to bind interaction partners. These structural insights could not be directly determined from the experimental data, highlighting the value of the ENSEMBLE approach.

Having probed the structure of the R region free state, interactions between the R region and NBD1 were examined. In NMR experiments using labeled R region and unlabeled NBD1, multiple segments of nonphosphorylated R region were found to bind to NBD1 with varying affinities. The interactions result in a dynamic complex in which individual R region segments bind and are released from NBD1 in a highly dynamic equilibrium. Phosphorylation relieves these interactions, with the exception of those near S768 in the R region when NBD1 is ATP-bound. Binding of nonphosphorylated R region to ΔF508 NBD1 is similar to wild-type NBD1, with largely similar residues involved. Adding unlabeled R region to 15N-labeled ATP-bound NBD1 causes extensive chemical shift changes and resonance broadening. Comparison of these spectra with those of a related construct for which resonance assignments are available suggests that multiple portions of NBD1 are affected upon binding of the R region. This may be due to the R region directly binding to more than one surface or to global conformational changes of the NBD1.

Properties of the SLC26A3 STAS domain and its binding to the phosphorylated R region were also explored. Comparison of NMR spectra indicates that the congenital chloride-losing diarrhea (CLD) causing mutation ΔY526/7 does not significantly alter STAS domain structural properties and that this mutation likely causes altered inter-domain contacts within SLC26A3. The phosphorylated R region interacts with the STAS domain using a primary site near S768 and a secondary site near L805. These interacting segments are similar to those that bind NBD1, indicating that direct competition between STAS and NBD1 may occur for phosphorylated R region binding. NBD1 and STAS interaction data allow the proposal of a model of dynamic R region
interactions in which various R region segments bind and are released from various interaction partners within CFTR and other proteins in a highly dynamic exchanging equilibrium that functions to integrate regulatory effects.

### 6.2 Future Directions

#### 6.2.1 Further structural characterization of the phosphorylated R region

Characteristics of the nonphosphorylated R region derived from a variety of structural probes have been used as inputs into the program ENSEMBLE, allowing visualization of a representation of the nonphosphorylated R region free state. Structural probes including secondary structural propensity (chemical shifts, Figure 2.5), flexibility ($R_2$ relaxation rates, Figure 3.1), and molecular size (PFG experiments, chapter 3) determined for the phosphorylated R region indicate that phosphorylation reduces helical propensity and alters R region structural contacts. To further characterize the phosphorylated R region free state, additional properties including molecular dimensions (SAXS), bond angle orientations (RDC), and inter-residue distances (PRE) will also be examined. This information can then be used as inputs into the program ENSEMBLE, as for the nonphosphorylated R region. Comparing representations of the nonphosphorylated and fully phosphorylated R region from ENSEMBLE will provide insight into the phosphorylation-dependent regulatory switches present in the R region. Possible effects of phosphorylation on the R region conformational ensemble may include increased compactness and stabilization of specific conformations that enhance binding interactions and exchange between various binding sites. Alternatively, reductions in flexibility could inhibit exchange between various binding sites. Changes in electrostatics may also modulate the overall molecular dimensions.

The effects of inhibitory versus stimulatory phosphorylation sites on R region conformations will also be probed by comparing these WT reference conformational ensembles to properties of R region states with only the inhibitory sites (737 and 768) or only the stimulatory sites (660, 670, 700, 712, 753, 795, and 813) phosphorylated (Wilkinson, Strong et al. 1997; Vais, Zhang et al. 2004). These R region states with only specific subsets of phosphorylation sites may lead to distinct conformational ensembles from the fully nonphosphorylated and fully phosphorylated states. Further
characterization of the R region free state and its modulation by partial or complete phosphorylation will inform both our understanding of CFTR regulation and protein regulation by disordered protein segments in general.

6.2.2 Mapping the binding interface of human R region and human NBD1

This thesis work examined the human R region residues involved in wild-type mouse NBD1 binding (Figure 4.2), probing the effects of ATP-binding and R region phosphorylation. The wild-type NBD1 residues involved were also examined (Figure 4.4). Examining the binding of human R region to human-derived NBD1 constructs would build upon this work. A human NBD1 construct encompassing residues 387-646, with RI residues 405-436 deleted, (hNBD1/ΔRI) has recently been crystallized (Atwell, Conners et al. 2007). This protein is more stable than the previous mouse NBD1 (mNBD1) protein fragments examined and $^{1}$H$^{15}$N spectra of this construct from our group show good resonance dispersion and more uniform peak intensities than the WT mouse NBD1, showing that this protein fragment is a good candidate for further study using NMR. Interpretation of R region binding to wild-type mNBD1 (Figure 4.4) was limited by the difficulty in transferring resonance assignments to this sample from those determined for mNBD1-RE in the presence of 3 suppressor mutations (G550E/R553M/R555K), as well as limited assignments in the original dataset (Kanelis, Hudson et al. 2009). Near-complete spectral assignment of the hNBD1/ΔRI resonances has been performed (Andrew Chong, personal communication). Also, the improved stability of this protein should permit the unlabeled R region to be titrated into hNBD1/ΔRI with multiple additions, allowing chemical shift changes to be tracked in $^{1}$H$^{15}$N spectra from the free to bound state. These experiments should give a much more complete characterization of the NBD1/R region interaction surface, while additional experiments using hNBD1 constructs with an intact RI should provide more information about the role of this second phospho-regulatory segment.

Work in this thesis examined the effects of PKA focusing on R region phosphorylation, however this should be expanded to include the effects of RI (S422) phosphorylation within the NBD1 primary sequence. For these experiments hNBD1 fragments encompassing residues 387 to 646 can be used with either the F494N or
F429S/F494N/Q637R solubilizing mutations that improve human NBD1 solubility and stability (Lewis, Zhao et al. 2005). These experiments will consider interactions between the R region and NBD1 where both regulatory segments are phosphorylated and in a system where both are derived from human CFTR. The phosphorylation status of S422 may modulate R region binding in addition to R region phosphorylation.

6.2.3 Comparison of wild-type and ΔF508 NBD1 binding to R region

Experiments presented in Figure 4.3 compare the binding of nonphosphorylated human R region to wild-type and ΔF508 ATP-bound mNBD1. Binding of the two is surprisingly similar considering that electrophysiological data finds altered channel activation of ΔF508 CFTR with phosphorylation (Wang, Zeltwanger et al. 2000). This suggests that the ΔF508 mutation may hinder the release of phosphorylated R region segments from NBD1 at a given level of PKA phosphorylation. In the context of mNBD1-RE, the ΔF508 mutation causes limited spectral changes in nonphosphorylated NBD1, however there are significant spectral differences in the phosphorylated state, suggesting altered dynamics of the RI and/or RE (Kanelis, Hudson et al. 2009). To begin addressing these differences, spectra of wild-type and ΔF508 in a hNBD1/ΔRI background can be compared to first examine potential structural or dynamic changes present within the core molecule, rather than as a consequence of altered phospho-regulatory region dynamics. Next, the binding of unlabeled nonphosphorylated and phosphorylated R region to labeled wild-type (Section 6.2.2) and ΔF508 hNBD1/ΔRI could be compared. If wild-type and ΔF508 hNBD1/ΔRI are similar in the free state but have altered R region binding, this would implicate changes in ΔF508 hNBD1 core properties that modulate R region binding. If both the free state structure and R region binding are similar between wild-type and ΔF508 hNBD1/ΔRI, altered RI interactions would be implicated in the ΔF508 defect.

These experiments would also be performed with labeled R region and unlabeled wild-type and ΔF508 NBD1/ΔRI. Experiments adding unlabeled ΔF508 mNBD1 to labeled R region (Figure 4.3) were very challenging, and only nonphosphorylated R region with ATP-bound ΔF508 mNBD1 could be probed because of limited ΔF508 mNBD1 stability. The improved solubility of hNBD1/ΔRI may allow the effects of phosphorylation and
ATP-binding to be probed for both wild-type and ΔF508 hNBD1/ΔRI as they were for wild-type mNBD1 (Figure 4.2). Any differences noted with the ΔF508 mutation may provide insight into the altered channel activation seen in ΔF508 CFTR (Wang, Zeltwanger et al. 2000).

6.2.4 Characterizing the dynamic R region – NBD1 complex

Having identified the R region and NBD1 residues involved in binding under various phosphorylation and ATP-binding conditions, further characterization of this interaction may facilitate a full molecular representation of this dynamic complex. Previous experiments have examined residues involved in the interaction, but experiments to specifically probe residues directly at the interaction interface will help to build a picture of the complex. Saturation transfer experiments could be used to detect the transfer of magnetization from NBD1 to R region residues in direct contact, rather than simply detecting residues involved in binding (Takahashi, Nakanishi et al. 2000). Additionally, PRE experiments in which broadening of resonances of one interaction partner is monitored upon interaction with the spin labeled other partner could be used to generate distance restraints. With the identification of direct interaction points between NBD1 and R region, the free state R region ensemble of conformations, and the NBD1 crystal structure, computational approaches may be used to dock the R region conformational ensemble onto NBD1, yielding a representation of the disordered complex.

6.2.5 R region binding to NBD2 and NBD1/NBD2 dimerization

While there is currently no evidence for direct interactions between the R region and NBD2, such an interaction is possible to further link phospho-regulatory events with nucleotide binding and hydrolysis in CFTR channel regulation. Even in the absence of direct NBD2 interactions, R region interactions with NBD1 may be modified in the presence of NBD2 because of NBD1/NBD2 dimerization. The influence of R region and RI phosphorylation status on NBD1/NBD2 dimerization has not yet been directly tested. Biochemical experiments such as pull-downs and Western blots on NBD1 and NBD2, monitoring the effects of the presence of R region and phosphorylation, may detect differences in dimer formation. To gain residue-specific information and also probe
potentially weaker interactions that would not be detected in biochemical experiments, NMR experiments can be performed to examine interactions between the R region and NBD2 in a similar manner to those with NBD1, including potential ATP-binding and phospho-modulation. Candidate interaction surfaces between the R region and NBD2 could involve further blocking of the NBD1/NBD2 dimerization interface or binding at an additional surface. The interactions of R region with NBD1 and NBD2 each in isolation could also be compared to those with an NBD1/NBD2 heterodimer. Approaches including using a catalytically inactive NBD2 E1471Q mutant (Moody, Millen et al. 2002; Vergani, Lockless et al. 2005) and/or high affinity ATP analogs such as P-ATP (Zhou, Wang et al. 2005) may help to stabilize the heterodimer. This work would provide a more complete structural characterization of the effects of phosphorylation and nucleotide binding regulatory events and enable more complete models of CFTR structure and regulation to be developed.

6.2.6 R region binding to N\textsubscript{t} and C\textsubscript{t} elbow helices

In addition to R region interactions with the NBDs, characterizing other candidate intramolecular CFTR/R region interactions would improve understanding of CFTR structure and function. The R region has been shown biochemically to interact with residues predicted to be just before the first transmembrane helix in the N-terminal region of CFTR (Naren, Cormet-Boyaka et al. 1999). These residues of CFTR may form an elbow helix (N\textsubscript{t} elbow helix) similar to that observed in the MsbA MSD-NBD homodimer (Ward, Reyes et al. 2007) (Figure 1.7). The interaction is mediated by negatively charged residues in the N\textsubscript{t} elbow helix, including a residue, D58N, whose mutation causes CF (http://www.genet.sickkids.on.ca/cftr/app). Mutations that prevent R region binding to the N\textsubscript{t} elbow helix also decrease P_{o} in the context of full length CFTR (Naren, Cormet-Boyaka et al. 1999; Fu, Ji et al. 2001).

There may also be an analogous C-terminal elbow helix (C\textsubscript{t} elbow helix) (Figure 1.7) before transmembrane helix 7. CFTR residues 846-853 align with residues in the MsbA elbow helix and are predicted to have helical propensity with AGADIR (Munoz and Serrano 1995) (Figure 6.1a). Representing these residues using a helical wheel diagram indicates that they may form an amphipathic helix that would be disrupted with
the CF-causing mutation R851L (http://www.genet.sickkids.on.ca/cftr/app) (Figure 6.1b). Residues in this putative helix are also included in the segment 838-859 implicated in R region binding (Gupta, Xie et al. 2004). Probing the structural features and phosphorylation dependence of interactions between the R region and the N_t and C_t putative elbow helices, and comparing WT with the D58N and R851L CF-causing mutants, may improve our understanding of the role of these segments in normal and disease states. This could be tested by synthesizing peptides corresponding to residues in the N_t and C_t elbow helices, titrating the peptides (constrained in micelles) into labeled R region samples, and monitoring any spectral changes. Interactions between the R region and these elbow helices, particularly if modulated by phosphorylation, could provide a mechanism to couple regulatory events in the CFTR cytoplasmic domains to structural rearrangements of the membrane spanning domains during channel gating.

**Figure 6.1 Properties of the putative amphipathic C_t elbow helix**

(a) Percent helix predicted using AGADIR (Munoz and Serrano 1995) for CFTR residues 845-859, contained in the CFTR segment 838-859 implicated in R region binding (Gupta, Xie et al. 2004). Alignment using ClustalX (Thompson, Gibson et al. 1997) of CFTR residues with *E. coli* MsbA residues is shown below the plot. The schematic indicates residues that form an elbow helix in the *E. coli* MsbA structure (3B5W). (b) Helical wheel representation of CFTR residues that align with the MsbA elbow helix. Hydrophobic (yellow), polar (green), and positively-charged (blue) residues are indicated. A red circle indicates the site of the R851L CF-causing mutation (http://www.genet.sickkids.on.ca/cftr/app). Residues 847-854 form an amphipathic helix.
6.2.7 Interactions of R region (817-838) negatively charged residues

Although the R region is highly charged, there is a non-uniform charge distribution throughout the primary sequence. In particular, positively charged residues primarily occur in the PKA consensus sites, while negatively charged residues are largely grouped in three clusters: residues 725-733 and 741-746 on either side of the inhibitory S737 PKA phosphorylation site (Wilkinson, Strong et al. 1997; Vais, Zhang et al. 2004) and residues 815-838 near the R region C-terminus (Figure 6.2). CFTR Δ817-838 opens without PKA and does not increase activity in response to phosphorylation (Xie, Adams et al. 2002), indicating the key role this C-terminal negative charge cluster plays in CFTR channel function, apparently decoupling normal channel regulatory events.

**Figure 6.2** Charge distribution in the R region

R region residues 654 to 838 are shown, with positively (blue) and negatively (red) charged residues indicated.

The CFTR surface(s) that interact with these negatively charged residues have not yet been determined. The CFTR primary sequence was examined for clusters of positively charged residues that are cytoplasmic, not within NBD1 (as residues 817-838 show only minor involvement in NBD1 interactions, Figure 4.2), and not within the N-terminal residues of CFTR encompassing 1-75 (interaction of this segment with the R region is mediated by negatively charged residues in the segment 1-75) (Naren, Cormet-Boyaka et al. 1999). A candidate positively charged CFTR segment was found within the last helix of ICD$_c$ (Figure 1.7), between transmembrane helix 12 and NBD2 in the primary sequence (Figure 6.3). CFTR residues 1155 to 1170 align with MsbA residues within the final helix of ICD$_c$. AGADIR predicts a helical conformation for residues from 1155 to 1168 (Figure 6.3a), that could form an amphipathic helix with one positively charged face (Figure 6.3b). C-terminal to this putative helix there are several prolines that would likely disrupt helical structure, however there are a series of other positively charged residues close in sequence so that CFTR residues 1155 to 1183 have a total net charge...
of +5 (Figure 6.3c). To probe whether this segment interacts with the C-terminal residues of the R region, a peptide corresponding to residues 1155-1168 could be synthesized, with additional flanking N- and C-terminal serine residues to improve solubility if required, and binding tested by the addition of peptide to $^{15}$N/$^{13}$C-labeled R region samples. Longer protein fragments encompassing residues 1155 to 1183 could also be expressed in *E. coli* and binding tested in a similar manner. These interactions may be required for proper coupling of cytoplasmic regulatory events to the channel pole through interactions between the R region and ICD$_c$ and explain the defects seen with CFTR Δ817-838.

![Figure 6.3](image)

**Figure 6.3** Candidate ICD$_c$ positively-charged surface for interactions with negatively-charged R region residues

(a) Percent helix predicted using AGADIR (Munoz and Serrano 1995) for CFTR residues 1155-1170 following transmembrane helix 12 in the primary sequence that likely form the last helix of ICD$_c$. Alignment using ClustalX (Thompson, Gibson et al. 1997) of CFTR residues with *E. coli* MsbA residues is indicated below the plot. (b) Helical wheel representation of CFTR residues with helical propensity predicted by AGADIR. Hydrophobic (yellow), polar (green), and positively charged (blue) residues are indicated. These residues form an amphipathic helix. (c) CFTR primary sequence following transmembrane helix 12 to before NBD2 with positively (blue) and negatively (red) charged residues indicated. Three prolines likely disrupt continuous helical structure over the entire segment, however the entire segment has a net local charge of +5.

### 6.2.8 R region interactions with SLC26A3 STAS domain

Preliminary work in this thesis has examined the interaction of phosphorylated R region and the SLC26A3 STAS domain by monitoring R region spectral changes with the addition of unlabeled STAS (Figure 5.2). These experiments could be repeated to look at binding of nonphosphorylated R region to STAS under the same experimental
conditions, as others have noted different phosphorylation dependence of this interaction (Ko, Zeng et al. 2004; Dorwart, Shcheynikov et al. 2008). Interaction experiments should also be repeated, examining R region binding to ΔY526/527 STAS to determine if this mutation alters R region binding, as a possible explanation of the CLD disease-causing mechanism. Using labeled STAS, NMR spectral assignment experiments could be performed. Spectral changes with the addition of unlabeled nonphosphorylated and phosphorylated R region would indicate the STAS residues involved in binding. STAS and NBD1 appear to compete for binding of the same phosphorylated R region segments (Figure 5.2), thus competition binding experiments will be useful to confirm this regulatory effect. Initial biochemical pull-downs can be followed by NMR titration experiments, beginning with unlabeled NBD1 added to labeled STAS + unlabeled R region samples and then other combinations of labeled/unlabeled interaction partners.

6.3 Conclusion

The work discussed in this thesis is one of the most complete structural characterizations of an intrinsically disordered protein segment, incorporating a variety of different structural probes. The ~180 residue intrinsically disordered R region segment is also one of the largest studied. The results provide novel residue-specific characterization of the CFTR R region and information on the phosphorylation-dependent binding to NBD1 and interactions with SLC26A3 STAS. This work reveals significant transient structure throughout the R region that is modified with phosphorylation and contributes to modulation of interactions with NBD1 and STAS in a dynamic complex. A model is proposed in which the R region acts as a regulatory hub, integrating interactions with a variety of partners to regulate channel function. This work provides the groundwork for future studies of the R region and its interactions to further probe properties of the free R region in various phosphorylation states and interactions with a variety of partners for a more detailed understanding of this fluctuating integrator of CFTR regulatory stimuli.
References


product binds to the second PDZ domain of the NHE3 kinase A regulatory
protein (E3KARP), potentially linking intestinal Cl-/HCO3- exchange to Na+/H+
exchange." Biochemistry 41(41): 12336-42.

Chichester, England, John Wiley & Sons, Ltd.


Lewis, H. A., X. Zhao, et al. (2005). "Impact of the deltaF508 mutation in first nucleotide-
binding domain of human cystic fibrosis transmembrane conductance regulator

structures representing the denatured state of the bovine acyl-coenzyme a

in human and characterization of SLC26A6, a candidate gene for pancreatic


Ma, J., J. Zhao, et al. (1997). "Function of the R domain in the cystic fibrosis
transmembrane conductance regulator chloride channel." J Biol Chem 272(44):
28133-41.


7804-5.


SH3 domain unfolded state suggest a compact ensemble with native-like and


