Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis

Innokentiy Maslennikov, Christian Klammt, Eunha Hwang, Georgia Kefala, Mizuki Okamura, Luis Esquivies, Karsten Mörs, Clemens Glaubitz, Wittek Kwiatkowski, Young Ho Jeon, and Senyon Choe

Edited by Robert M. Stroud, University of California, San Francisco, CA, and approved April 30, 2010 (received for review February 11, 2010)

For Biomolecular Magnetic Resonance, Institute for Biophysical Chemistry, Goethe University, Frankfurt 60438, Germany

Membrane domain structures of three classes of histidine kinase receptors (HKRs). The ArcB and QseC TM domains are both two-helical motifs, whereas the KdpD TM domain comprises a four-helical bundle with shorter second and third helices. The interhelical distances (up to 12 Å) reveal weak interactions within the TM domains of all three receptors. Determined consecutively within 8 months, these structures offer insight into the abundant and underrepresented in the Protein Data Bank class of 2-4 TM crossers and demonstrate the efficiency of our CF combinatorial dual-labeling strategy, which can be applied to solve MP structures in high numbers and at a high speed. Our results greatly expand the current knowledge of HKR structure, opening the doors to studies on their widespread and pharmaceutically important bacterial signaling mechanism.

Histidine kinase receptors (HKRs) are part of a two-component system, in which an HKR in the bacterial inner membrane transmits a signal to a response regulator located in the cytoplasm (1). This signaling system constitutes the predominant signal transduction mechanism by which bacteria interact with their environment (1), however, many aspects of its signaling mechanism are unknown. Based on the part of the receptor used to sense environmental stimuli, HKRs are grouped into three structural classes. Class 1, the largest class, is characterized by the presence of an extracytoplasmic sensory domain that senses external stimuli and transmitting the signal across the membrane. Class 2 HKRs lack an apparent extracytoplasmic domain and the stimuli-sensing region is believed to be in the membrane domain itself. Class 3 is characterized by a cytoplasmic sensory domain. For this study, we selected QseC (class 1, quorum sensor), ArcB (class 2, aerobic respiratory control sensor), and KdpD (class 3, membrane sensor) as representatives from each of the three classes. HKRs are multidomain proteins with diverse structure and topologies (Fig. 1): QseC has two transmembrane (TM) helices with an intervening 130 amino acid periplasmic sensor domain, ArcB has two TM helices connected by a short periplasmic loop, and KdpD has four TM helices with short interhelical loops. The cytoplasmic regions of HKRs comprise a dimerization domain and a kinase domain, several structures of which have been reported (2–9), including that of ArcB. There are also several reported periplasmic sensor domain structures (10–13). However, due to the highly flexible multidomain nature of HKRs, there is still no structure of a full-length receptor, nor has any structural information on an HKR membrane domain been reported to date. Such a structure is essential to understand the mechanistic aspect of signal transduction.

Despite impressive progress in the structure determination of membrane proteins (MPs) by X-ray crystallography and NMR


The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The NMR-derived data (atomic coordinates, chemical shifts, and restraints) have been deposited in the Protein Data Bank, www.pdb.org, under accession codes 2KSD [ArcB(1-115)], 2KSE [QseC(1-185)], and 2KSF [KdpD(397-502)].

1I.M. and C.K. contributed equally to this work.

2To whom correspondence may be addressed. E-mail: yjeon@kbsi.re.kr or choe@salk.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1001656107/-/DCSupplemental.

PNAS | June 15, 2010 | vol. 107 | no. 24 | 10902–10907
spectroscopy in recent years (see reviews in refs. 14 and 15), only about 250 structures of unique integral MPs have been determined, representing less than 1% of known protein structures (16). In addition to problems with expression, solubilization, and purification of integral MPs, X-ray and NMR methods are hampered with inherent technical difficulties. Diffraction quality crystals of integral MPs are very difficult to obtain because solubilized protein-detergent complexes do not usually form ordered crystal lattices. Furthermore, the mobility of TM helices leads to strong broadening of the signals in NMR spectra and presents problems with spectra quality, resonance assignment, and detection of long-range interactions, all of which are necessary to calculate the structure of a protein by NMR. Spin label-based paramagnetic relaxation enhancement (PRE) approaches have been used to address the paucity of long-distance constraints associated with α-helical integral MPs (17, 18). However, the cost of isotope labeling by in vivo heterologous expression in cells of both prokaryotic and eukaryotic origins is often prohibitive for NMR structural studies of even well-expressed MPs.

We addressed the aforementioned technical difficulties, developing a strategy for a cost-efficient structure determination of MPs that has allowed us to quickly solve the backbone structures of three HKR membrane domains by NMR. The chosen structures represent the three classes of this widespread bacterial TM signaling system and belong to the abundant group of 2–4 TM crossers, which is poorly represented in the Protein Data Bank (PDB). By studying the domains in isolation, we present the structural insights into the signal-transducing membrane domains of these receptors. The results demonstrate the efficiency of the combinatorial dual-labeling (CDL) strategy. Thus, based on synergy of fast cell-free (CF) expression and rapid NMR analysis, CDL opens up possibilities for accelerated structural analysis of MPs.

Results

Structures of MPs Produced in CF System Are Comparable to Those from Escherichia coli. To synthesize membrane domains of selected histidine kinases we used the precipitating CF (p-CF) expression mode (19), in which a protein is produced as a precipitate that is subsequently solubilized by a non-denaturing detergent (a comprehensive description is given in SI Materials and Methods). The p-CF mode is extremely useful because it allows NMR studies of MPs without purification (Fig. S1). Because all of the CF reaction components are soluble, they remain in a supernatant and are easily removable after reaction by pellet wash. As a result, the target MP can be expressed without affinity tags that might affect its structure or stability. Although reminiscent of bacterial inclusion bodies, the MP precipitate can be solubilized with a mild lipid-like detergent (19) unlike true inclusion bodies, which require strong denaturing conditions (20). Therefore, it is likely that the CF precipitate has at least a partially folded structure.

To verify this, we investigated the underlying structure of the precipitates by magic angle spinning solid-state NMR (MAS SS-NMR). Uniformly 13C-labeled ArcB(1-115) and KdpD(397-502) precipitates were measured directly after their production in the p-CF reaction. By measuring the deviation of the 13C chemical shifts away from their typical values for a random coil peptide, we could differentiate between helical, beta, and random coil secondary structures (21) in the precipitates. In these two samples, all of the observed 13C-13C cross-peaks for alanine and valine residues fell within the regions of the 13C-13C correlation spectra typical for α-helices (21) (Fig. 24 and Fig. S2). The 13C-13C cross-peaks for valine and alanine residues that reside in the loop regions are probably broadened beyond the detectability limits in the MAS SS-NMR spectra. Moreover, the spectrum of the ArcB(1-115) precipitate is very similar to the spectrum of the ArcB(1-115) sample lyophilized after solubilization with a detergent (Fig. 24, red contours). The presence of secondary structure in the precipitate was confirmed by solution NMR measurements of proton-deuteron (H-D) exchange following detergent solubilization. The results show that, in ArcB(1-115), 11 out of 16 valines and 5 out of 7 alanine are located in protected and therefore likely TM helical regions. Thus MAS SS-NMR analysis of chemical shifts together with solution NMR data on the exchange of the labile backbone protons in the precipitate (Figs. S3 and S4) have unambiguous interpretation: The TM helices of ArcB (1-115) and KdpD(397-502) existed as secondary structure elements before precipitation in a CF reaction.

To further validate the p-CF expression system, we compared it with the standard E. coli expression in terms of sample quality and structure, as well as time and cost efficiency (Fig. 3 and Table S1). ArcB(1-115) was expressed and purified using both approaches. The E. coli system required five consecutive days, beginning with the transformation and growth of bacteria in minimal media and then extraction and purification of the protein from the cell membrane, in order to start the first NMR measurement. In contrast, using the p-CF synthesis, the first NMR measurement was
and improved later by the high selectivity of isotope labeling in CF synthesis and the efficiency of modern NMR heteronuclear techniques (23). Several approaches were built upon this method to perform site-selective screening (24) and to accelerate the resonance assignment by using dual combinatorial labeling (25–27).

Simplified approach utilizing combinatorial selective $^{15}$N labeling was proposed for determination of the amino acid types for $^{1}$H$^{15}$N, $^{13}$NH resonances (28). Combinatorial selective labeling approaches (reviewed in refs. 29 and 30) use a different subset of labeled amino acids in each sample in order to reduce the number of samples required to achieve the assignment as compared to one-by-one selective labeling. Most of the existing combinatorial approaches use labeling schemes that are sequence blind and optimized for globular proteins (25, 26, 28). These work for proteins that give good signal dispersion and good overall spectra quality, but are impractical for MPs.

Expanding on the signal originally proposed by Trbovic (27) to resolve ambiguities in the NMR assignment of CF-expressed MPs, we created a CDL strategy that is applicable to MPs (details are described in Materials and Methods). At the core of the CDL strategy is the individual optimization of the labeling scheme for each protein sequence and each selected set of amino acids for isotopic labeling. To derive these sequence-dependent schemes, we have developed the software MCCl (Monte Carlo Combinatorial Labeling, available at http://sbl.salk.edu/combipro), which calculates the optimal labeling combination for a given protein sequence with a defined number of samples using the Monte Carlo approach. As a result, the CDL method ensures that a minimal number of samples are needed to assign the type of the first and the second amino acid in every pair in a sequence, meanwhile minimizing the spectral complexity of each sample.

The CDL strategy was used for the design of $^{15}$N$^{13}$C-labeling schemes for both KdpD(397-502) (Fig. 4B) and QseC(1-185) (Table S2), consisting of six and seven labeled samples, respectively. Within 1 day after spectra collection, 27% and 22% of $^{1}$H$^{15}$N cross-peaks for KdpD(397-502) and QseC(1-185), respectively, were unambiguously assigned. Additionally, the amino acid type was defined for 100% and 74% of the $^{1}$H$^{15}$N cross-peaks for KdpD(397-502) and QseC(1-185), respectively. The incomplete rapid assignment provides evenly distributed multiple starting points for traditional sequential assignment throughout

---

**Fig. 3.** Comparison of performance of the CF system with the standard *E. coli* system. SDS-PAGE shows marker (M); CF reaction mixture (RM) before reaction at 0 h (1); CF RM after ArcB(1-115) expression at 15 h (2); precipitate after ArcB(1-115) p-CF (3); *E. coli* (EC) expressed ArcB(1-115) after extraction, purification, Tag cleavage, size-exclusion chromatography (SEC), detergent exchange on Q-sepharose, and concentration (4); arrow indicates ArcB(1-115).

**Fig. 4.** The CDL strategy for the “point-directed assignment” of NMR spectra. (A) The $^{1}$H$^{15}$N cross-peak in HSQC will appear only if the second residue in a pair is $^{13}$N labeled (second box tagged 1). The $^{1}$H$^{15}$N cross-peak in the HSQC spectrum and the $^{1}$H$^{15}$N$^{13}$C cross-peak in the HNCO spectrum will both appear only if the peptide group is double $^{15}$N labeled (third box tagged 2). (B) Dual $^{1}$H$^{15}$N$^{13}$C combinatorial selective labeling scheme designed specifically for backbone assignment of KdpD(397-502) (see details in SI Text). (C) Assignment of $^{1}$H$^{15}$N$^{13}$C combinatorial cross-peaks of KdpD(397-502) using a combinatorial scheme of sequential $^{15}$N labeling, presented in panel B. The overlays of $^{1}$H$^{15}$N$^{13}$C$^{15}$N$^{13}$C-TROSY-HNCO (red contours) and $^{1}$H$^{15}$N projection of TROSY-HNCO (blue contours) spectra are shown for each sample (I–VI). Absence of a cross-peak (tag “0”), a cross-peak present in TROSY only (tag “1”), and cross-peaks present in both the TROSY and the HNCO spectra (tag “2”) in each combinatorially labeled sample determine the code (sequence of the tags) for every cross-peak A, B, and C in a uniformly labeled sample. The scheme-designed code which is identical with the code determined from the recorded spectra defines a pair of amino acids for the $^{1}$H$^{15}$N, $^{13}$NH, and $^{13}$C resonances.
the sequence. Starting from and building upon the results of the CDL-derived assignment, the standard sequential assignment procedure (31, 32) was tremendously accelerated, yielding an assignment for 100% of KdpD (397-502) and 76% of QseC (1-185) backbone resonances in 2–3 weeks. ArcB (1-115) resonances (96% of the backbone, 88% of C\text{\textbeta}, and most of H\text{\textalpha} and H\text{\textbeta}) were assigned using the standard sequential assignment protocol. The rapid assignment of backbone resonances enabled us to analyze the secondary structure based on the chemical shift index and the hydrogen exchange data (summarized in Fig. S4) and to proceed with de novo NMR structure determination using the structural constraints collected for assigned backbone atoms.

**HKR TM Domain Structures.** Having verified that the p-CF system yields a folded protein and completed the NMR resonance assignments, we collected the NOE and PRE data required to calculate the structures of ArcB (1-115), QseC (1-185), and KdpD (397-502) (see Materials and Methods). The folds depicted in Figs. 1B and 5 show the helical nature of the TM domains from these three proteins whose calculated structures have a crystallographic equivalent resolution (33) in the range of 2.7–3.2 Å (Table 1).

The structures of ArcB (1-115) and QseC (1-185) (Figs. 1B and 5A and B) are both two-helical motifs whose helices have the appropriate length (30–32 Å) to cross the lipid bilayer. The large periplasmic signaling domain of QseC was not included in the structure calculations and so its TM domain is simply composed of two antiparallel weakly interacting α-helices with a crossing angle of 157 ± 4°. Similarly, the connecting loop in ArcB (1-115) is not well defined and its TM domain comprises two antiparallel α-helices with a crossing angle of 142 ± 6° and a distance of 11.1 Å at the closest point between the helices (all helical packing parameters are summarized in Table S3). In comparison, the crossing angle between two helices of HTR-II Transducer in complex with sensory Rhodopsin (34) is 169° and the distance between the helices is 10 Å, whereas for two tightly packed helices in dimeric human glycophorin A, the distance is just 6.4 Å (35). Prolines at position 67 of ArcB and positions 166 and 167 in QseC disrupt the helical hydrogen bond patterns (see the \textsuperscript{1}H exchange data for ArcB (1-115) in Fig. S4) and create kinks of 22 ± 2° [ArcB (1-115)], 22 ± 5° and 24 ± 4° [QseC (1-185)] in the second helix, adding local flexibility to the helices and increasing the interhelical distances near the periplasmic side of the membrane, thus additionally weakening the helix–helix interactions.

The TM domain of KdpD comprises a four-helix bundle (Figs. 1B and 5C, and Table S3), in which the second and third helices are relatively short (23–24 Å) and loosely packed with the crossing angle of −165 ± 6° and the interhelical distance of ~9.4 Å. The second helix interacts with the first and third helices only. The first and the fourth helix show the crossing angle of −157 ± 4°. These two helices weakly interact only near their cytoplasmic ends and this is the only consistent interaction involving the first helix, which causes the whole bundle to be packed rather loosely. It is possible that the loose helical packing observed in all three structures could be an artifact of the detergent solubilization or particular choice of HKR truncation. Nonetheless, the TM structures lead us to surmise that the loose helical packing provides an inherent flexibility in the TM domains and that this is perhaps essential to the mechanism of signal transduction across the membrane. The idea needs further verification including computational dynamics simulation studies.

**Discussion**

We have taken a deconstructionist approach to tackling the problems of MP structure determination. Because the size and mobility of MPs are limiting factors in NMR analysis, the study of individual domains has obvious advantages. Although domain-wise structure determination is nothing new to crystallography, the impetus is usually to remove offending domains (those that inhibit crystallization), which in the case of MPs is most often the membrane domain itself, or conversely engineering artificial soluble protein domains into MPs in order to obtain diffraction quality MP crystals (36). NMR techniques, on the other hand, are more amenable to the structural study of flexible membrane regions: ArcB (1-115), residues 20-83; QseC (1-185), residues 10-38; and 156-185 (KdpD (397-502)). Backbones are shown for the stable regions: ArcB (1-115), residues 20-83; QseC (1-185), residues 10-38; and 156-185 and KdpD (397-502), residues 397-502. Consecutive TM helices are colored in the following order: green, blue, orange, and coral. Structures on the right are rotated 90° vertically.

Fig. 5. Twenty superimposed structures of the TM domains of (A) ArcB (1-115), (B) QseC (1-185), and (C) KdpD (397-502). Backbones are shown for the stable regions: ArcB (1-115), residues 20-83; QseC (1-185), residues 10-38; and 156-185 and KdpD (397-502), residues 397-502. Consecutive TM helices are colored in the following order: green, blue, orange, and coral. Structures on the right are rotated 90° vertically.

Maslennikov et al. PNAS vol. 107 no. 24 10905

PNAS | June 15, 2010 | vol. 107 | no. 24 | 10905
domains. However, this does not mean that such domains necessarily retain their native states, and the interpretation of the structures needs to be made with this caveat in mind.

The hallmark feature of these three TM domain structures appears to be their loose helical packing. The packing of TM \( \alpha \)-helices is related to protein function and can be rigid, as in the case of channel pores like KcsA (37), ionotropic receptors like nACHR (38), the glutamate receptor channel (39), and tightly packed multihelical proteins like membrane respiratory enzymes (40), or flexible, as observed in the case of many metabotropic membrane receptors like G protein-coupled receptors (36) and kinase receptors. The majority of the solved structures of integral MPs (>97%) represents proteins that actively or passively transport something (a molecule, ion, proton, or electron) across the biological membrane (channels and transporters) and proteins that tightly bind another molecule to catalyze a reaction (oxidases, ATPases, intramembrane proteases, etc.). The metabotropic membrane receptors are still a much underrepresented family in the PDB. Their primary role in a cell is to transmit signals through the membrane. Therefore, they do not require a well-defined conformational state of the TM domain that would be needed, for example, for coordinating transported ions or molecules. On the contrary, in order to transmit a signal, the TM domain must act as a conformational switch, an action which can be supported by its intrinsic mobility (41).

Furthermore, although it is known that HKRs act as dimers in the membrane, the fact that all of the structures are monomers can be explained by the absence of the cytoplasmic dimerization domain in our constructs. The dimerization of HKRs is unlikely mediated by TM domains as reflected by their loose helical packing. Further studies need to be done to address the significance of the loose helical packing of the transmembrane signaling mechanism of HKRs.

Concluding Remarks

The three structures presented in this study offer a glimpse into the abundant class of 2-4 TM crossers, underrepresented in the PDB, and provide an important inroad toward understanding the mechanistic aspects of the conformation-driven signal transduction process. The CDL strategy, founded in the synergy between the CF and NMR methods, opens up new possibilities for fast determination of backbone structures of MPs, especially those recalcitrant to crystallization. It is also complementary with the recently proposed approach by Raman et al. (42) for fast determination of backbone structure of large globular proteins. Backbone structures determined quickly by the CDL strategy would provide excellent starting points for high-throughput modeling of a large number of classes of integral MPs and further structure-function prediction.

Materials and Methods

**Protein Expression, Characterization, and NMR Sample Preparation.** ArcB(1-115) was expressed in a throbmin-cleavable N-terminal His\(_6\) tag in E. coli BL21 DE3 cells (Invitrogen) and purified as described in SI Materials and Methods. ArcB(1-115), QseC(1-185), and KdpD(397-502) were expressed as p-CF in the absence of detergents (19). Washed precipitate was solubilized in 300 \( \mu \)l 5% (w/v) 1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)] (Avanti Polar Lipids; Anatrace), 20 mM Mes-BisTris pH 5.5 [ArcB(1-115)] or pH 5.7 [QseC(1-185)] and KdpD(397-502) were expressed in the CF system were system were solubilized at 45' and 37°C on a Bruker AVANCE 700 MHz spectrometer (Salt). Details of NMR experiments are described in SI Materials and Methods.

**CDL Strategy.** The CDL strategy comprises four steps: (i) designing the combinatorial labeling scheme with the MCLL program; (ii) parallel expression of 6-7 CDL samples using the p-CF expression system (0.5—1.0 mL of reaction mixture for each CDL sample) and solubilization in the same buffer to eliminate any differences in cross peak positions; (iii) short measurement of \( ^{1}H-^{15}N\)-TROSY-HSQC and 2D-HNCO spectra (about 0.5—1.5 h per spectrum) for each CDL sample, all the samples for the combinatorial assignment of a particular protein were measured in only 1—2 days, depending on the concentration of the protein; (iv) analysis of the spectra and assignment of \( ^{1}H-^{15}N\)-cross-peaks using the CARA program (44).

**Structure Calculation and Analysis.** The \( ^{13}C\) chemical shift deviations from random coil values were used to define backbone torsion angle restraints (45). The conformational distance restraints were derived from the backbone torsion angle restraints of NOE cross peaks measured in 3D \( ^{15}N\)-resolved TROSY-\( ^{1}H-^{13}C\)-NOESY (mixing time 120 ms). The long-range distance constraints were derived by analysis of the PRE effect as described (18). An interactive procedure, which included structure calculation by the CYANA program (46) followed by the distance constraints refinement, was used to calculate the backbone spatial structures of ArcB(1-115), QseC(1-185), and KdpD(397-502). The summary of the structural statistics of the structures is presented in Table 1. The 20 conformers with the lowest target function of the last CYANA calculation cycle were energy-minimized using the CNS program (47). The helical packing parameters, such as interhelical crossing angles, interhelical distances, and helical kinks, were subsequently analyzed with the Helix Packing Pair program (48).

**Point-Directed Assignment.** The point-directed assignment is based on selective semi-isotopic (\( ^{15}N \) and \( ^{13}C \)) labeling of two types of amino acids, located consecutively in a protein sequence. For every pair of residues in which the first amino acid is labeled with \( ^{13}C\) and the second amino acid is labeled with \( ^{15}N\), cross-peaks in both the \( ^{1}H-^{15}N-^{15}N\)-HSQC and the HNCCO spectra appear (tag "2" in Fig. 4A). If the second amino acid in a pair is labeled with \( ^{15}N\) and the first amino acid is not labeled with \( ^{13}C\), there will only be a cross-peak in the HNCCO spectrum (tag "1" in Fig. 4A). For pairs in which the second amino acid is not labeled with \( ^{15}N\), there will be no cross-peaks in either spectrum (tag "0" in Fig. 4A). Thus, by analyzing the presence and absence of cross-peaks in the \( ^{15}N-^{15}N\)-HSQC and the HNCCO spectra recorded from a selected labeled sample, one can define the type of amino acids for backbone resonances.

### Table 1. Summary of NMR data and statistics for the calculated sets of 20 lowest energy structures of ArcB(1-115), QseC(1-185), and KdpD(397-502)

<table>
<thead>
<tr>
<th></th>
<th>ArcB(1-115)</th>
<th>QseC(1-185)</th>
<th>KdpD(397-502)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR distance constraints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE*</td>
<td>72</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>PRE</td>
<td>291</td>
<td>295</td>
<td>845</td>
</tr>
<tr>
<td>Hydrogen bonds</td>
<td>31</td>
<td>28</td>
<td>56</td>
</tr>
<tr>
<td>NMR dihedral angle constraints</td>
<td>37</td>
<td>42</td>
<td>72</td>
</tr>
<tr>
<td>Phi</td>
<td>37</td>
<td>42</td>
<td>72</td>
</tr>
<tr>
<td>Psi</td>
<td>37</td>
<td>42</td>
<td>72</td>
</tr>
</tbody>
</table>

Violations (mean and SD):
- Distance constraints, Å: 0.80 ± 0.01, 1.18 ± 0.12, 0.72 ± 0.02
- Dihedral angle constraints, °: 10.3 ± 1.70, 10.3 ± 0.40, 5.70 ± 0.51
- Max. distance violation, Å: 0.17 ± 0.01, 0.27 ± 0.03, 0.22 ± 0.02
- Max. dihedral angle violation, °: 1.74 ± 0.01, 2.12 ± 0.40, 0.81 ± 0.01
- Backbone rms deviation, Å: 1.45 ± 0.45, 2.35 ± 0.85, 1.61 ± 0.48
- Average pairwise in the set: 1.41 ± 0.46, 2.18 ± 0.86, 1.56 ± 0.49
- Equivalent resolution (Å): 2.9, 3.2, 2.7

*Interresidue sequential (\( i−j \) = 1) constraints only.
*Applied in TM helical regions.
*Calculated by PROCHECK program (33).
participating in these cross-peaks and bind to a particular amino acid pair. If a pair is unique in the sequence, an exact assignment of the \(^1\)H, \(^{13}\)N, and \(^{15}\)C resonances to the residues that comprise that pair is instantly made. Numerically, each pair of residues in a given sample is specified by a tag depending on its labeling combination, as explained in Figure 4A. Therefore, in a series of combinatorially labeled samples a pair of residues is defined by a series of tags, that is, a code which is directly related to the presence or absence of cross-peaks in both spectra for each sample. The scheme-designed code which is identical with the code derived from the recorded spectra defines a pair of amino acids for the \(^1\)H, \(^{13}\)N, and \(^{15}\)C resonances. If the code is unique, i.e., there is only one pair in the sequence it corresponds to, the assignment of the \(^1\)H and \(^{15}\)N resonances to the second residue, as well as the assignment of the \(^{13}\)C resonance to the first residue of the pair, is made. This simple analysis provides an unambiguous assignment for ~30–40% of the backbone \(^1\)H, \(^{13}\)N, and \(^{15}\)C resonances and defines the amino acid type for the rest of the backbone \(^1\)H, \(^{13}\)N, and \(^{15}\)C resonances, limiting the number of their possible positions in a sequence to as few as 2–4.

**CDL Assignment Process.** The CDL assignment process is demonstrated for three KdpD(397-502) cross-peaks in Figure 4C. Here, an \(^{13}\)C–\(^1\)N cross-peak for residue C is present in the TROSY spectra of samples I, III, IV, and V and in the HN plane of the HNCO spectrum of sample IV, therefore, its code is 101210 (the digit place corresponds to sample number). This code is unique and corresponds to the 1H47-\(^{13}\)Ala482 pair in the sequence, which gives us an unambiguous assignment for Ala482. Cross-peak B has the code 021102 and was assigned to three possible Ala-Val pairs in the sequence (Val414/72/483). Cross-peak A has the code 011101 and was assigned to nine possible pairs, with Val as the second amino acid in every pair and Arg, Val, or Thr, which are not labeled by \(^{13}\)C, as the first amino acid.

**Acknowledgments.** This work was supported by National Institutes of Health (GM074929), Incheon Free Economic Zone, and World-Class University program (S.C.), and Korea Membrane Protein Initiative and 21C Frontier Microbial Genomics program (Y.H.J.). C.K. was partly supported by Deutsche Forschung Gemeinschaft Research Fellowship.