J Biomol NMR (2011) 51:115–122
DOI 10.1007/s10858-011-9546-9

ARTICLE

Visualizing the principal component of ^{1}H,^{15}N-HSQC NMR spectral changes that reflect protein structural or functional properties: application to troponin C

Ian M. Robertson · Robert F. Boyko · Brian D. Sykes

Received: 11 May 2011 / Accepted: 13 July 2011

Abstract Laboratories often repeatedly determine the structure of a given protein under a variety of conditions, mutations, modifications, or in a number of states. This approach can be cumbersome and tedious. Given then a database of structures, identifiers, and corresponding ^{1}H,^{15}N-HSQC NMR spectra for homologous proteins, we investigated whether structural information could be ascertained for a new homolog solely from its ^{1}H,^{15}N-HSQC NMR spectrum. We addressed this question with two different approaches. First, we used a semi-automated approach with the program, ORBplus. ORBplus looks for patterns in the chemical shifts and correlates these commonalities to the explicit property of interest. ORBplus ranks resonances based on consistency of the magnitude and direction of the chemical shifts within the database, and the chemical shift correlation of the unknown protein with the database. ORBplus visualizes the results by a histogram and a vector diagram, and provides residue specific predictions on structural similarities with the database. The second method we used was partial least squares (PLS), which is a multivariate statistical technique used to correlate response and predictor variables. We investigated the ability of these methods to predict the tertiary structure of the contractile regulatory protein troponin C. Troponin C undergoes a closed-to-open conformational change, which is coupled to its function in muscle. We found that both ORBplus and PLS were able to identify patterns in the ^{1}H,^{15}N-HSQC NMR data from different states of troponin C that correlated to its conformation.

Keywords Troponin C · L48Q · Structure prediction · ORBplus · Principal component · Partial least squares

Introduction

The determination of protein structures by NMR spectroscopy is a valuable tool used by scientists that span a wide array of disciplines. Laboratories often repeatedly determine the structure of a protein under a variety of conditions, mutations, modifications, or in a number of ligand-bound states; however, this can be tedious. For example, medicinal chemists are interested in protein-drug structures to aid in the design of novel drugs, but determination of many similar structures with different drugs can be too slow to be effective in the discovery pipeline. As a result, several other approaches have been adopted over the years to help rapidly characterize protein and protein–ligand structures. Chemical shift mapping utilizes the basic hypothesis that if a ligand binds a target, the chemical environment surrounding residues lining the interaction surface will be changed as a consequence of binding (for reviews on this and other NMR screening methods see (Pellecchia et al. 2002; Stockman and Dalvit 2002)). Typically, researchers use ^{15}N-labeled protein to rapidly scan these interactions in order to localize a binding site and quantitate ligand stoichiometry and affinity. The benefit of this approach is it can readily distinguish whether the target and ligand are interacting and may be able to predict a binding surface. The downside is that amide chemical shifts are sensitive to a variety of sources including structure, dynamics, hydrogen bonding, and solvent accessibility; therefore, chemical shift perturbations induced by

Electronic supplementary material The online version of this article (doi:10.1007/s10858-011-9546-9) contains supplementary material, which is available to authorized users.

I. M. Robertson · R. F. Boyko · B. D. Sykes (✉)
Department of Biochemistry, University of Alberta, Edmonton, AB T5K 2W7, Canada
e-mail: brian.sykes@ualberta.ca

ligand binding may be remote from its binding site. For example, chemical shift perturbations of amide resonances may stem from an overall conformational or stability change in the target upon binding a ligand. In this case, simply chemical shift mapping will not be effective at localizing the binding site of a ligand. On the other hand, if a series of compounds induce a similar conformational change in the target, then this perceived limitation of amide chemical shift mapping can be utilized to the user's advantage. Given a database of structures, identifiers (i.e. target conformation), and corresponding $^{1}H,^{15}N$-HSQC NMR spectra for a set of homologous proteins, it is interesting to ask: can structural information be ascertained for a new homolog solely from its $^{1}H,^{15}N$-HSQC NMR spectrum?

We address this possibility in the context of the cardiac regulatory protein troponin C. Cardiac troponin C (cTnC) is the Ca^{2+}-binding subunit of the heterotrimeric troponin complex. The two other proteins in the complex are: troponin I (cTnI), the inhibitory subunit, and troponin T (cTnT), the subunit that tethers the complex to the muscle fiber. cTnC has two high-affinity metal binding sites in its C-terminal lobe (cCTnC) that remain bound to Ca^{2+} or Mg^{2+} during contraction and relaxation; and one lower affinity Ca^{2+} binding site in its N-terminal lobe (cNTnC). Muscle contraction is triggered when Ca^{2+} binds to cNTnC and induces a conformational change from a 'closed' to a 'partially open' state so that cNTnC can bind cTnI. At low Ca^{2+} levels, cNTnC is in its apo state and cTnI is bound to actin, preventing contraction from occurring. When Ca^{2+} binds cNTnC the switch region of cTnI ($cTnI_{147–163}$) interacts with and stabilizes the 'fully open' state of cNTnC and its inhibition of contraction is abrogated. The structures of a number of these physiological states of cNTnC have been solved by NMR spectroscopy and X-ray crystallography. The apo and Ca^{2+} bound states of cNTnC and cTnC were solved by NMR spectroscopy (Sia et al. 1997; Spyracopoulos et al. 1997). The structure of cNTnC-Ca^{2+} in complex with $cTnI_{147–163}$ was also solved by NMR (Li et al. 1999). Later, the X-ray structure of the core troponin complex (cTnC-3Ca^{2+}-$cTnI_{31–210}$-$cTnT_{183–288}$) was solved (Takeda et al. 2003). In addition to these natural states of cTnC, a number of structures have been solved for cNTnC-Ca^{2+}-drug (Hoffman and Sykes 2009; Li et al. 2000; Igarashi et al. 2005) and cNTnC-Ca^{2+}-cTnI-drug (Oleszczuk et al. 2010; Robertson et al. 2010; Wang et al. 2002) complexes.

The tertiary structure of troponin C can be defined by the AB and CD inter-helical angles. While it is convenient to discuss the structure of cNTnC in terms of a single static conformation, it is important to emphasize that the structures are dynamic and are more correctly viewed as an ensemble. For example, the Ca^{2+}-saturated state of cNTnC is not actually 'partially open' but rather is in a dynamic equilibrium between open and closed states (Paakkonen et al. 2000; McKay et al. 2000; Eichmueller and Skrynnikov 2007). When $cTnI_{147–163}$ binds cNTnC it stabilizes its open state (see Fig. 1a). The corresponding $^{1}H,^{15}N$-HSQC is shown in Fig. 1b—which highlights the large chemical shift perturbations induced by the addition of $cTnI_{147–163}$. The side chain of residue Leu48 of cNTnC makes intramolecular hydrophobic contacts with A23, F20, and F27 which stabilizes the closed conformation of the AB inter-helical angle

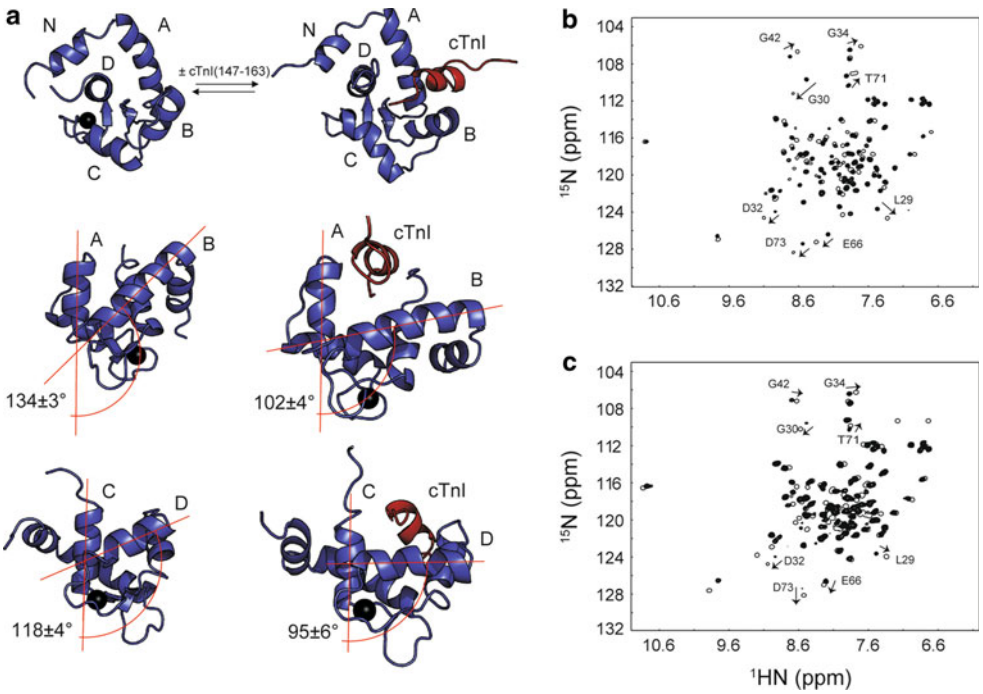

Fig. 1 The closed-to-open transition of cardiac troponin C upon binding troponin I. **a** The structures of cNTnC and the cNTnC-$cTnI_{147–163}$ complex are shown. The five helices of cNTnC are labeled (N, A, B, C, D). The inter-helical angles, AB and CD, for the two states of cNTnC are shown below the corresponding structures. As the interhelical angles decrease towards 90°, cNTnC is more open. **b** The overlay of $^{1}H,^{15}N$-HSQC spectra from cNTnC (*filled contours*) and the cNTnC-$cTnI_{147–163}$ complex (*open contours*). **c** The overlay of $^{1}H,^{15}N$-HSQC spectra from cNTnC (*filled contours*) and cNTnC(L48Q) (*open contours*). Several of the residues that are significantly perturbed by $cTnI_{147–163}$ and the L48Q mutation are labeled

observed for cNTnC-Ca^{2+} (Sia et al. 1997). Tikunova and Davis postulated that the mutation of this residue to the hydrophilic residue, glutamine, should disrupt these hydrophobic interactions and thus stabilize a more open form of cNTnC (Tikunova and Davis 2004). Their functional studies on this mutation indicated that cNTnC(L48Q) was more active than cNTnC, probably due to its more open conformation. In Fig. 1c we overlay the $^{1}H,^{15}N$-HSQC spectra of cNTnC-Ca^{2+} with cNTnC(L48Q)-Ca^{2+}. Many of the same residues that are perturbed by $cTnI_{147–163}$ are also shifted by this leucine to glutamine mutation. This protein thus serves as a query protein for the data mining approaches we investigate herein to answer the straightforward but subtle question: *can we confidently define the inter-helical angles that define the troponin C's tertiary structure from chemical shift data in lieu of complete structure determination*?

We used two different approaches to look for patterns in $^{1}H,^{15}N$ HSQC spectra: a semi-automated program ORBplus and the multivariate statistical analysis—partial least squares (PLS). ORBplus looks for patterns in the chemical shifts and correlates these commonalities to the explicit property of interest. The program ranks resonances based on consistency of the magnitude and direction of the chemical shifts within the database, and the chemical shift correlation of the unknown protein with the database and visualizes the results by a histogram and a vector diagram, and provides residue specific predictions on structural similarities with the database. The PLS approach is a statistical technique that is used to maximize covariance between responses and predictors of corresponding observable variables. PLS and other multivariate analyses, such as principal component analysis, are popular in the analysis of large databases such as identifying interesting motions in a molecular dynamics trajectory (Caves et al. 1998; Ichiye and Karplus 1991; Mu et al. 2005; Kitao et al. 1998) and metabolic foot-printing in metabolomics (Stoyanova and Brown 2001; Lindon et al. 2001).

Study case results

Semi-automated analysis with ORBplus

ORBplus is a semi-automated tool for the identification of patterns within $^{1}H,^{15}N$-HSQC spectra that correspond to structural or functional properties of a protein.[1] ORBplus follows a similar semi-automated approach and design utilized by the NMR assignment tool, Smartnotebook (Slupsky et al. 2003), and the NMR chemical shift predictor tool, ORB (Gronwald et al. 1997). ORBplus works by comparing the assigned $^{1}H,^{15}N$-HSQC NMR spectrum of a query protein of unknown properties with a database of $^{1}H,^{15}N$-HSQC NMR spectra from proteins with corresponding known properties. The feature in question can be, but is not limited to, a structural property. ORBplus takes NMRView (Johnson and Blevins 1994) peaklist files or Biological Magnetic Resonance Bank (BMRB) formatted files as input for both the database $^{1}H,^{15}N$-HSQC data and the query chemical shift data. It requires at least two datasets that represent two different states (i.e. open and closed or active and inactive). ORBplus treats each chemical shift in the $^{1}H,^{15}N$-HSQCs as coordinates and connects them via a vector. In the case where many datasets exist for a given state, the $^{1}H,^{15}N$-HSQC data are averaged and a vector is drawn to the averaged $^{1}H,^{15}N$-HSQC data point by default; however, the user may decide to compare the query dataset to just one database dataset as well. ORBplus makes projections by automatically ranking residues as good predictors by four criteria: (1) the magnitude of chemical shift in the database files, (2) similarity of chemical shift within the database files, (3) magnitude of chemical shift for the query dataset, and (4) the direction of chemical shift for the query data in comparison with the database datasets. ORBplus also has the option for the user to manually select residues of interest that may not have been automatically selected by ORBplus. Once the predictor residues have been chosen, ORBplus predicts the query data's properties by projecting the query chemical shift onto the average database vector and comparing the fraction length of the query shift to the average. For example, in the case where we are considering inter-helical angles, if the projection of the query data is 75% the length of the average vector, then ORBplus predicts for that residue, the inter-helical angle of the query protein is 75% that of the average angle in the database.

We provided ORBplus with $^{1}H,^{15}N$-HSQC data for corresponding cNTnC structures with known AB and CD interhelical angles (see Table 1). First we looked to see how ORBplus would perform given $^{1}H,^{15}N$-HSQC of the cNTnC-Ca^{2+}-$cTnI_{147–163}$ complex (unless otherwise stated all forms of cNTnC mentioned will assume Ca^{2+} is bound). Although ORBplus will automatically pick the top ten residues to predict the unknown query protein's properties, we wanted to probe for specific changes in the two inter-helical angles. To do this, we chose to separately follow residues from, or directly adjacent to, Ca^{2+}-binding site I (residues 27–40), and Ca^{2+}-binding site II (residues 64–74), because it has been shown that these regions experience large chemical shift perturbations when cNTnC's conformation changes (Li et al. 1999). Using these residues, ORBplus predicted an AB inter-helical angle of 104° and a CD angle of 90°. These results are within the error of those determined experimentally: 102 ± 4° and 95 ± 6° (NMR)

[1] The ORBplus program and extensive documentation is freely available at http://www.bionmr.ualberta.ca/bds/software/orbplus/.

Table 1 Inter-helical angles of cNTnC structures

cNTnC structures	PDB code	Interhelical angles	
		AB[a]	CD[b]
cNTnC(apo) (NMR)	1SPY	140 ± 3°	127 ± 5°
cNTnC-Ca^{2+} (NMR)	1AP4	134 ± 3°	118 ± 4°
cNTnC(acys)-Ca^{2+} (NMR)	2CTN	142 ± 3°	109 ± 4°
cNTnC-Ca^{2+}-$cTnI_{147–163}$ (NMR)	1MXL	102 ± 4°	95 ± 6°
cTnC(acys)-3Ca^{2+}-$cTnI_{31–210}$-$cTnT_{183–288}$ (X-ray)	1J1D	104°	96°
cNTnC(F77W,V82A)-Ca^{2+} (NMR)	2JXL	108 ± 5°	95 ± 4°
cTnC(acys)-3Ca^{2+}-3bepridil (X-ray)	1DTL	93°	89°
cNTnC-Ca^{2+}-$cTnI_{147–163}$-bepridil (NMR)	1LXF	121 ± 4°	85 ± 4°
cNTnC(acys)-Ca^{2+}-2TFP (X-ray)	1WRK	107°	104°
cNTnC(acys)-Ca^{2+}-W7 (NMR)	2KFX	114 ± 3°	86 ± 2°
cNTnC(acys)-Ca^{2+}-$cTnI_{147–163}$-W7 (NMR)	2KRD	112 ± 5°	63 ± 6°
cNTnC(acys)-Ca^{2+}-$cTnI_{144–163}$-dfbp-o (NMR)	2L1R	96 ± 6°	89 ± 6°

The smaller the interhelical angle, the more open the cNTnC structure is (see Fig. 1a). The reference closed state of cNTnC used by ORBplus (Fig. 2) is highlighted in grey

Angles calculated with the program Interhlx (K. Yap, University of Toronto)

[a] AB interhelical angles were determined by defining the C helix as residues 17–26, and the D helix was defined as residues 40–46 for all structures

[b] CD interhelical angles were determined by defining the C helix as residues 54–62, and the D helix was defined as residues 75–83 for all structures

or 104° and 96° (X-ray).[2] The vector diagrams for residues in site I and site II are shown in Fig. 2.

Once we established that ORBplus could predict the inter-helical angles of cNTnC when in complex with $cTnI_{147–163}$ we tested it on cNTnC(L48Q). Leu48 lies at the end of the B helix of cNTnC and makes several hydrophobic interactions with residues along the A helix thereby stabilizing the closed conformation of cNTnC, even in the presence of Ca^{2+} (see Table 1). The mutation of leu48 to a glutamine is thought to disrupt these crucial hydrophobic interactions: the strong NOEs that were observed by Sia et al. between the terminal methyls of Ala23 and Leu48 (Sia et al. 1997) are not observed between the terminal NH_2 of Gln48 and the β-methyl of Ala23 (Supplementary Figure 1). The top ten automatically chosen residues by ORBplus are displayed in Table 2. For the same reasons described above, we manually selected residues in sites I and II for the calculation of the AB and CD interhelical angles. ORBplus predicted that the AB inter-helical angle of cNTnC(L48Q) is slightly more open than cNTnC(wt) (124° vs. 134 ± 3°), whereas the CD interhelical angle is practically unchanged (113° vs. 118 ± 4°). These results are as anticipated and illustrate the efficacy of ORBplus at predicting even subtle perturbations in the conformation of cNTnC.

[2] Angles calculated with the program Interhlx (K. Yap, University of Toronto).

Multivariate analysis with partial least squares

Multivariate analysis is a useful method used to extract a fundamental property from a large amount of data, and recently it has been used to look at 2D-NMR data. Sakurai and Goto used principal component analysis (PCA) to correlate the pH-dependent conformation of β-lactoglobulin with ^{1}H,^{15}N-HSQC spectra (Sakurai and Goto 2007). They were able to identify a key subset of residues that underwent chemical shift changes as a function of pH—an observation they were able to link to several structural transitions. Jaumot et al. utilized multivariate curve resolution alternating least squares analysis to understand the underlying mechanism of a chemical reaction monitored by ^{1}H,^{15}N-HSQC spectroscopy (Jaumot et al. 2004). Multivariate analysis has also been used in other aspects of structural biology, for example it resolved the crucial intermolecular features of diubiquitin involved in stabilizing the different lysine-linked ubiquitin chains (Fushman and Walker 2010).

We subjected the same ^{1}H,^{15}N-HSQC NMR chemical shift data used in ORBplus to a PLS analysis using the statistical software Simca P+ v12.0.1 (Umetrics, Umeå, Sweden; Eriksson et al. 2001). PLS was used to maximize the correlation between the NMR data and the conformation of cNTnC. We designated the chemical shift data as the X variables and the inter-helical angle data as the Y variables and allowed Simca to use the standard defaults for mean centering and scaling. All residues were given a

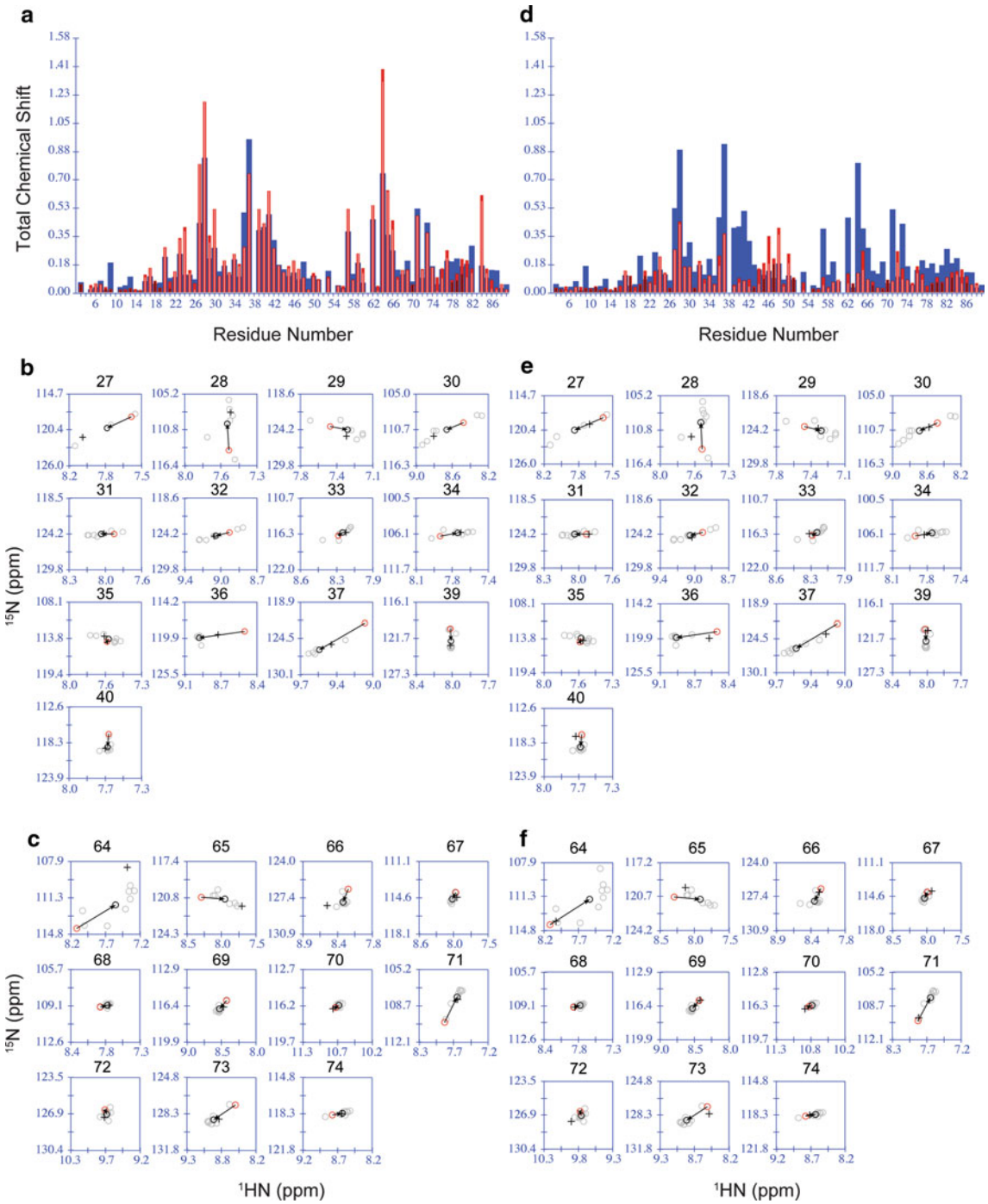

Fig. 2 ORBplus chemical shift histogram and spectral plots displaying the prediction of the conformation of (**a**–**c**) the cNTnC-cTnI$_{147–163}$ complex and (**d**–**f**) for cNTnC(L48Q). **a**, **d** Histograms compare the magnitude of chemical shift changes of the query protein (*red*) and the average chemical shift of the database proteins (*blue*) from the reference closed state (see Table 1). The histograms also compare the direction of chemical shift change with the average chemical shift change (*white inner bar* indicates the same direction and *blue inner bar* indicates opposite direction—see ORBplus documentation for details). Spectral perturbations for residues (**b**, **e**) in site I and (**c**, **f**) site II are shown. The *red circles* represent the chemical shift of the reference closed state (see Table 1), the *light grey circles* represent the chemical shifts for the database proteins, and the *black circles* the average chemical shift of the database proteins. The cross is the chemical shift of the query protein

corresponding angle: residues 3–52 were associated with the AB inter-helical angle and residues 53–89 were given CD inter-helical angles [cNTnC(L48Q) data only included chemical shifts since we do not have experimentally derived inter-helical angles]. The scatter plot is shown in Fig. 3a. The plot indicates that there are two groupings:

Table 2 Top ten residues as indicators of the conformation of cNTnC

ORBplus	PLS (^{1}HN chemical shifts)	PLS (^{15}N chemical shifts)
S37 (site I)	F74 (site II)	A22
V28 (site I)	G34 (site I)	F77
I36 (site I)	G68 (site II)	E59
V64 (site II)	E32 (site I)	D73 (site II)
F27 (site I)	E59	S37 (site I)
E66 (site II)	L29 (site I)	T38 (site I)
L41	S37 (site I)	E40 (site I)
T71 (site II)	L41	V28 (site I)
D73 (site II)	D62	E66 (site II)
D65 (site II)	A31 (site I)	E32 (site I)

Residues in site I or site II are identified. The first column lists the ten most predictive residues of conformation of cNTnC(L48Q) as automatically chosen by ORBplus (using the default parameters) based on the correlation of the chemical shift magnitude and direction of cNTnC(L48Q) with the database chemical shifts. The second and third columns are the top ten residues from the VIP plot for the ^{1}HN (Fig. 3b) and ^{15}N (Fig. 3c) chemical shifts

one including the closed states of cNTnC (2CTN, 1AP4) and the other more dispersed various open states of cNTnC (2JXL, 1WRK, 1MXL, 2KFX, 1DTL, 1LXF, 2KRD, and 2L1R).

The variable influence on projection (VIP) plots (Fig. 3b, c) can be used to highlight the most important residues for fitting the PLS model; the higher the VIP score the more the residue contributes to the model. The results from the ^{1}HN and ^{15}N nuclei clearly indicate that the most important chemical shifts reside in the two loops (sites I and II) of cNTnC (see Table 2 for a list of the top 10 residues). To compare the predictions from PLS with those from ORBplus, we used the PLS model to predict the angles of cNTnC-cTnI$_{147–163}$ and cNTnC(L48Q). We only looked at predictions for residues in site I and site II, as was done in ORBplus. PLS predicted an AB inter-helical angle of 105 ± 1° and a CD inter-helical angle of 89 ± 0.1°. These results are similar to those determined by Interhlx [AB: 102 ± 4° and CD: 95 ± 6° (NMR); AB: 104° and CD: 96° (X-ray)]. The angles predicted by the model for cNTnC(L48Q) were 123 ± 2° for the AB inter-helical angle and 105 ± 0.4° for the CD inter-helical angle. The AB inter-helical angle is very close between the ORBplus and PLS predictions; however, the CD inter-helical angle prediction has a variance between the two methods of ~8° (ORBplus = 113°; PLS = 105 ± 0.4°).

Discussion

The use of chemical shifts to predict protein structure has been employed by others (Shen et al. 2008, 2009; Wishart et al. 2008); however, these methods all require the assignment of ^{13}C-labeled protein to predict tertiary structure. In a previously published study, Biekofsky et al. used ab initio methods to correlate ^{15}N chemical shift differences to Ca^{2+}-coordination and protein conformation in troponin C and other EF-hand proteins (Biekofsky et al. 2004). In this work, detailed analysis of solely ^{1}H,^{15}N-HSQC NMR spectra from a variety of conformational states of cNTnC was able to provide insight into the tertiary structure of cNTnC(L48Q). We used two methods to analyze the NMR data: a semi-automated program, ORBplus; and the multivariate statistical method, PLS. Both programs predicted that cNTnC(L48Q) was more open than cNTnC at the AB inter-helical interface, but differed in their prediction of the effect of the mutation on the CD inter-helical angle. ORBplus predicted only a minor perturbation, while the PLS model predicted a slightly more open angle.

The quality factors ($R^2Y = 0.509$ and $Q^2 = 0.303$) for the PLS model were slightly less than that expected for a biological sample ($R^2Y = 0.5$ and $Q^2 = 0.4$) (Eriksson et al. 2001), which illustrates the model's limitations in its fit and predictability. There are several explanations for PLS's mediocre predictability: slight variations in sample conditions (pH, protein concentration, protein–ligand ratios), lack of consistency between the conditions used in the structure determination and in the ^{1}H,^{15}N-HSQC experiments (X-ray structures vs. NMR data), ^{1}H,^{15}N-HSQC data were acquired with just cNTnC whereas X-ray structures often included full length cTnC, missing data in the ^{1}H,^{15}N-HSQC spectra (TFP binds to cNTnC in the intermediate exchange regime, and some peaks were not able to be followed throughout the titration), and perhaps most importantly, the conformational model is simplistic. Clearly some of the chemical shift changes could result from proximity to the different ligands/mutations and *not* solely from a conformational change, such as ring current effects induced by aromatic rings present on the ligands. In

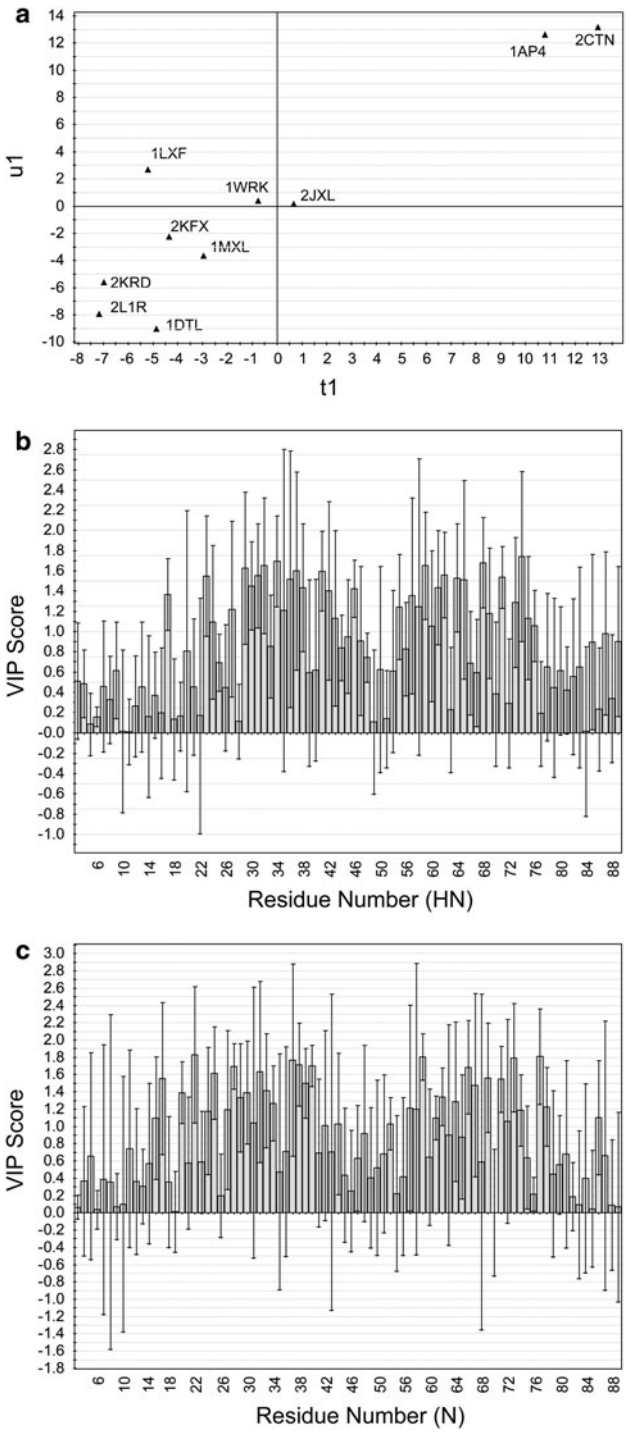

Fig. 3 PLS scatter and VIP plots of the $^{1}H,^{15}N$-HSQC data for the cNTnC structures. **a** u_1/t_1 scatter plot shows the correlation between inter-helical angles and chemical shifts. The cNTnC structures are labeled by their protein databank (PDB) codes (see Table 1). The PLS analysis used a one component model with $R^2Y = 0.509$ and $Q^2 = 0.303$. **b**, **c** Variable influence on projection (VIP) scores for $^{1}H,^{15}N$-HSQC data versus the sequence of cNTnC

fact, this phenomenon has been exploited by McCoy and Wyss to predict the binding site of a ligand on a protein with the program Jsurf (McCoy and Wyss 2002). However, in the case of troponin C, trying to predict a binding site by ring current effects using amide chemical shifts has led to erroneous binding site predictions, because the large chemical shifts induced by ligand binding are largely dominated by the opening and closing of troponin C (Robertson et al. in press). It may be that the more straightforward approach of ORBplus, augmented by the interactive GUI and human intuition, is a better method when dealing with noisy data. One may envision a two pronged approach: (1) use PLS to identify the residues that correlate to the predictor variable, and (2) switch to ORBplus to analyze the data in a semi-automated way—selecting the most important residues chosen by PLS—to predict structure or function. ORBplus is not meant to compete with other multivariate methods of analysis, but to provide the user with a detailed visualization of the spectral changes that most reflect a change in a given property (such as activity, stability, structure, etc.) for the protein in question. In our system, it led to the immediate demonstration that the closed-to-open transition could be different for the AB and CD hinges depending upon the mutation or ligand.

There are a number of obvious advantages to using these types of computational approaches in predicting protein structure. For example, understanding how a large number of mutations perturb protein structure without repeating the cumbersome task of structure determination. Another application may be in pharmaceutical research; multivariate analysis is already being utilized by medicinal chemists for lead optimization (for a review see Gabrielsson et al. 2002). The ability of multivariate techniques to predict the conformational change induced when a protein binds a lead compound may be very useful, particularly if the compound of interest is designed to modulate the protein's tertiary structure in a specific manner. These models need not replace structure determination either: the conformational predictions could be used as restraints in a structure calculation to aid in refinement. It is important to note that although this work has emphasized the role $^{1}H,^{15}N$-HSQC analysis can play on predicting protein structure it does not have to be limited to this; in any situation where NMR chemical shifts and an observable (i.e. enzyme kinetics, ligand binding) are correlated, this type of investigation can be applied.

Acknowledgments The authors are thankful to Professor Michael Regnier for suggesting the L48Q project that resulted in this paper and discussions; to Melissa Crane for cloning, expressing, and purifying cNTnC(L48Q); Dr. David Broadhurst and Stacey Reinke for insight into the use of multivariate analysis; and Drs. Monica Li, Ryan

Hoffman, Marta Oleszczuk, and Olivier Julien for providing data for analysis and for suggestions made on the development of ORBplus. I.M.R. is the recipient of an AHFMR studentship. Supported by the Canadian Institutes of Health Research.

References

Biekofsky RR, Turjanski AG, Estrin DA, Feeney J, Pastore A (2004) Ab initio study of NMR N-15 chemical shift differences induced by Ca2+ binding to EF-hand proteins. Biochemistry 43:6554–6564

Caves LSD, Evanseck JD, Karplus M (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. Protein Sci 7:649–666

Eichmueller C, Skrynnikov NR (2007) Observation of mu s time-scale protein dynamics in the presence of Ln(3+) stop ions: application to the N-terminal domain of cardiac troponin C. J Biomol NMR 37:79–95

Eriksson LJE, Kettaneh-Wold N, Trygg J, Wikstrom C, Wold S (2001) Multi- and megavariate data analysis. Part 1: principles and applications. Umetrics Academy, Umeå

Fushman D, Walker O (2010) Exploring the linkage dependence of polyubiquitin conformations using molecular modeling. J Mol Biol 395:803–814

Gabrielsson J, Lindberg NO, Lundstedt T (2002) Multivariate methods in pharmaceutical applications. J Chemometr 16:141–160

Gronwald W, Boyko RF, Sonnichsen FD, Wishart DS, Sykes BD (1997) ORB, a homology-based program for the prediction of protein NMR chemical shifts. J Biomol NMR 10:165–179

Hoffman RMB, Sykes BD (2009) Structure of the inhibitor W7 bound to the regulatory domain of cardiac troponin C. Biochemistry 48:5541–5552

Ichiye T, Karplus M (1991) Collective motions in proteins—a covariance analysis of atomic fluctuations in molecular-dynamics and normal mode simulations. Proteins-Struct Funct Genetics 11:205–217

Igarashi T, Takeda S, Mori H (2005) Crystal structure of the N-terminal domain of human cardiac troponin C in complex with a calcium-sensitizer; trifluoperazine. J Mol Cell Cardiol 39:1016

Jaumot J, Marchan V, Gargallo R, Grandas A, Tauler R (2004) Multivariate curve resolution applied to the analysis and resolution of two-dimensional [H-1, N-15] NMR reaction spectra. Anal Chem 76:7094–7101

Johnson BA, Blevins RA (1994) NMR view—a computer-program for the visualization and analysis of NMR Data. J Biomol NMR 4:603–614

Kitao A, Hayward S, GO N (1998) Energy landscape of a native protein: jumping-among-minima model. Proteins-Struct Funct Genetics 33:496–517

Li MX, Spyracopoulos L, Sykes BD (1999) Binding of cardiac troponin-I147–163 induces a structural opening in human cardiac troponin-C. Biochemistry 38:8289–8298

Li Y, Love ML, Putkey JA, Cohen C (2000) Bepridil opens the regulatory N-terminal lobe of cardiac troponin C. Proc Natl Acad Sci USA 97:5140–5145

Lindon JC, Holmes E, Nicholson JK (2001) Pattern recognition methods and applications in biomedical magnetic resonance. Prog Nucl Magn Reson Spectrosc 39:1–40

Mccoy MA, Wyss DF (2002) Spatial localization of ligand binding sites from electron current density surfaces calculated from NMR chemical shift perturbations. J Am Chem Soc 124:11758–11763

Mckay RT, Saltibus LF, Li MX, Sykes BD (2000) Energetics of the induced structural change in a Ca2+ regulatory protein: Ca2+ and troponin I peptide binding to the E41A mutant of the N-domain of skeletal troponin C. Biochemistry 39:12731–12738

Mu YG, Nguyen PH, Stock G (2005) Energy landscape of a small peptide revealed by dihedral angle principal component analysis. Proteins-Struct Funct Bioinfo 58:45–52

Oleszczuk M, Robertson IM, Li MX, Sykes BD (2010) Solution structure of the regulatory domain of human cardiac troponin C in complex with the switch region of cardiac troponin I and W7: the basis of W7 as an inhibitor of cardiac muscle contraction. J Mol Cell Cardiol 48:925–933

Paakkonen K, Sorsa T, Drakenberg T, Pollesello P, Tilgmann C, Permi P, Heikkinen S, Kilpelainen I, Annila A (2000) Conformations of the regulatory domain of cardiac troponin C examined by residual dipolar couplings. Eur J Biochem 267:6665–6672

Pellecchia M, Sem DS, Wuthrich K (2002) NMR in drug discovery. Nat Rev Drug Discov 1:211–219

Robertson IM, Sun YB, Li MX, Sykes BD (2010) A structural and functional perspective into the mechanism of Ca2+-sensitizers that target the cardiac troponin complex. J Mol Cell Cardiol 49:1031–1041

Robertson IM, Pineda-Sanabria S, Sykes BD (in press) Approaches to protein-ligand structure determination by NMR spectroscopy: applications in drug binding to the cardiac regulatory protein troponin C. Biophys Struct Count Threats Chall

Sakurai K, Goto YJ (2007) Principal component analysis of the pH-dependent conformational transitions of bovine beta-lactoglobulin monitored by heteronuclear NMR. Proc Natl Acad Sci USA 104:15346–15351

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu GH, Eletsky A, Wu YB, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 105:4685–4690

Shen Y, Vernon R, Baker D, Bax A (2009) De novo protein structure generation from incomplete chemical shift assignments. J Biomol NMR 43:63–78

Sia SK, Li MX, Spyracopoulos L, Gagne SM, Liu W, Putkey JA, Sykes BD (1997) Structure of cardiac muscle troponin C unexpectedly reveals a closed regulatory domain. J Biol Chem 272:18216–18221

Slupsky CM, Boyko RF, Booth VK, Sykes BD (2003) Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments. J Biomol NMR 27:313–321

Spyracopoulos L, Li MX, Sia SK, Gagne SM, Chandra M, Solaro RJ, Sykes BD (1997) Calcium-induced structural transition in the regulatory domain of human cardiac troponin C. Biochemistry 36:12138–12146

Stockman BJ, Dalvit C (2002) NMR screening techniques in drug discovery and drug design. Prog Nucl Magn Reson Spectrosc 41:187–231

Stoyanova R, Brown TR (2001) NMR spectral quantitation by principal component analysis. NMR Biomed 14:271–277

Takeda S, Yamashita A, Maeda K, Maeda Y (2003) Structure of the core domain of human cardiac troponin in the Ca(2+)-saturated form. Nature 424:35–41

Tikunova SB, Davis JP (2004) Designing calcium-sensitizing mutations in the regulatory domain of cardiac troponin C. J Biol Chem 279:35341–35352

Wang X, Li MX, Sykes BD (2002) Structure of the regulatory N-domain of human cardiac troponin C in complex with human cardiac troponin I147–163 and bepridil. J Biol Chem 277:31124–31133

Wishart DS, Arndt D, Berjanskii M, Tang P, Zhou J, Lin G (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Res 36:W496–W502