Chemical tagging with t-butyl and trimethylsilyl groups for measuring intermolecular NOEs in a large protein-ligand complex


Abstract: Intermolecular 1H-1H nuclear Overhauser effects (NOE) present a powerful tool to assess contacts between proteins and binding partners, but are difficult to identify for complexes of high molecular weight. We show that intermolecular NOEs can readily be observed following chemical labeling with t-butyl or trimethylsilyl groups. Proteins can be furnished with t-butyl or TMS groups site-specifically using genetically encoded unnatural amino acids or by chemical modification of single cysteine residues. No isotope labeling is required. The approach is demonstrated with the 95 kDa complex between tetrameric *E. coli* single-stranded DNA binding protein (SSB) and single-stranded DNA.

The short-range nature of the nuclear Overhauser effect (NOE) makes it an excellent tool for obtaining intermolecular distance restraints in structural studies of protein-ligand complexes in solution. Increasing molecular mass of a complex, however, not only increases the complexity of its 1H-NMR spectrum, but also the 1H-NMR line widths due to increased 1H-relaxation rates, making the assignment of specific intermolecular 1H-1H NOEs challenging. The conventional approach is to improve spectral resolution by labeling with stable isotopes ([13]C and [15]N) and using more involved NMR pulse sequences such as half-filters and three-dimensional NMR spectra.[3] Unavoidably, however, long pulse sequences entail significant signal loss due to relaxation. To combat the enhanced relaxation rates associated with high molecular weight systems, perdeuterated samples are frequently prepared to limit the dipolar relaxation pathways of the remaining 1H-NMR spins,[4] but perdeuteration can change the properties of the protein[5] and the dilution of 1H-NMR spins decreases the chances of observing intermolecular 1H-1H contacts in the protein-ligand complex.

Here we show that intermolecular NOEs can be observed for proteins of high molecular weight at natural isotopic abundance by the use of t-butyl or trimethylsilyl (TMS) groups as chemical tags. These groups can be site-specifically installed in proteins either by using genetically encoded unnatural amino acids or by chemical modification of a single cysteine residue. We have previously shown that O-t-butyl-tyrosine (Tby) can easily be incorporated into proteins in response to an amber stop codon and that, even in the context of proteins with >100 kDa molecular weight, solvent-exposed t-butyl groups produce a narrow and intense 1H-NMR signal that can readily be observed.[6]

The TMS group equally features a singlet of nine equivalent 1H spins, but no unnatural amino acid with a TMS group has been reported that can be incorporated into proteins in response to a stop codon. TMS groups are attractive because of their chemical shift of about 0 ppm, where signal overlap with other resonances in the 1H-NMR spectrum is minimal. To introduce a TMS group, we labelled a cysteine residue with a TMS tag (Figure 1).

![Figure 1](image_url)

**Figure 1.** Chemical structure of the trimethylsilyl (TMS) tag used in the present work. The tag reacts with proteins by alkylation of cysteine residues.

We applied these labeling strategies to the complex between single-stranded DNA binding protein (SSB) from *Escherichia coli* and single-stranded DNA (ssDNA) presented by a 67-mer thymidine oligomer, d(T67). *E. coli* SSB is a homotetramer and the complex with d(T67) has a molecular mass of 95 kDa. A co-crystal structure of *E. coli* SSB in complex with two dC3 oligonucleotides has been published (PDB ID 1EYG).[7] The crystals were obtained following removal of the flexible C-terminal domain by chymotryptic digestion and show the ssDNA wrapping around the OB domains of the SSB tetramer. It is known from biochemical experiments that, in the presence of high salt (>200 mM NaCl), a SSB tetramer occludes a stretch of 65 nucleotides of ssDNA, indicating that each OB domain in the tetramer binds to ssDNA.[8-10] Although the crystal structure is missing electron density for parts of the dC3 oligomers, it suggests how ssDNA may wrap around SSB in this binding mode (referred to as (SSB)4). Unexpectedly, however, the ssDNA binding sites of different SSB monomers accommodate the ssDNA in different ways, with different orientations of the nucleotide bases relative to the protein (Figure 2).[7]

[a] Ms. S. Jabar, Dr. Y. Wang, Prof. G. Otting
Research School of Chemistry
Australian National University
Canberra, ACT 2601 (Australia)
E-mail: gottfried.otting@anu.edu.au

[b] Dr. L. A. Adams, Dr. L. Aurelio, Prof. B. Graham
Monash Institute of Pharmaceutical Sciences
Monash University
Parkville, VIC 3052 (Australia)
E-mail: bim.graham@monash.edu

† The first two authors contributed equally. Supporting information for this article is given via a link at the end of the document.
In the present work, we used a construct with C-terminal His-tag (referred to as EcSSB) and introduced a t-buty1 group by substituting Tyr97 by O-t-buty1-tyrosine-(phenyl-3,5-d2) (62-Tby) or O-t-buty1-tyrosine-(t-buty1-13C3) (13C-Tby). The Tby residue was incorporated by expressing EcSSB by cell-free synthesis, using a PETMCSI vector with an amber stop codon at position 97 and adding purified p-cyanophenylalanine-RNA synthetase, cognate tRNA\textsubscript{Ec}C\textsubscript{TMS}, and Tby. To tag EcSSB with the TMS tag of Figure 1, we expressed the mutant Thr98Cys \textit{in vivo} and incubated the purified protein with a five-fold excess of tag overnight (see the Supporting Information for details).

The 13C-HSQC spectrum of EcSSB Tyr97Tby showed a single cross-peak at a 1H chemical shift of 1.16 ppm with a full line width at half height of about 11 Hz (Figure 3). This chemical shift is within 0.2 ppm of the chemical shift of the free Tby amino acid. In the NOEY spectrum, the singlet of the t-buty1 group produced intense cross-peaks with resonances at about 6.8 and 7.0 ppm, which can be attributed to the aromatic ring protons of Tby (Figure 4).

To record intermolecular NOEs with ssDNA, we added 67-mer ssDNA in equimolar ratio. Initially, we used two different 67-mers, dT\textsubscript{67} and dT\textsubscript{67} with nucleotides 2, 3, 4 and 6 substituted by guanosine (dTG\textsubscript{2}G\textsubscript{67}). The 13C-HSQC spectra of either complex showed a single peak at a 1H chemical shift of 0.09 ppm, suggesting that the guanines near the 5’-end of the ssDNA do not
approach the t-buty1 group in any of the OB domains (Figure 3). Furthermore, the single cross-peak indicates that each OB domain binds ssDNA in the same way or that exchange between different conformations is fast. Complex formation was accompanied by an increase in line width of the 1H-NMR signal to 22 Hz. In the NOESY spectra, cross-peaks with the t-buty1 group were readily resolved because most NOEs between protein protons were broadened beyond detection (Figure 4). In the presence of ssDNA, the NOESY spectrum showed additional cross-peaks with the t-buty1 group, which must be intermolecular NOEs with dT67 (at 2.1 and 4.1 ppm, compare Figure 4A and B). Based on the chemical shifts, these NOEs are with sugar protons of the ssDNA. A peak at 1.6 ppm is more ambiguous, as it could be an NOE cross-peak with methyl groups of dT67 or an artefact associated with the intense diagonal peak. If it were with one or more thymidine methyl groups, a cross-peak with the nearby 6H proton of the base would be expected too. We therefore investigated the possibility that intermolecular NOEs with base protons of dT67 overlap with the intramolecular NOEs with the aromatic ring protons of the Tby residue. This was done by preparing a second sample, where the c-protons, which are ortho to the O-t-buty1 group, were replaced by deuterons. The corresponding NOESY spectrum showed the absence of the intra-residue cross-peak at 6.8 ppm as expected, and the cross-peak at 7.0 ppm was significantly weakened, indicating that it is with the β-protons of the Tby residue (Figure 4C). These results suggest that dT67 projects its phosphodiester backbone towards the residue at position 97, implying that contacts with the bottom of the ssDNA binding groove are formed by nucleotide bases rather than the backbone. The crystal structure 1EYG [7] is ambiguous in this regard (Figure 2).

Due to baseline distortions near the diagonal peak of the t-buty1 group, it was difficult to probe for intermolecular NOEs with methyl groups of dT67, which would be a sensitive indicator of contacts with the nucleotide bases. To check for such NOEs and probe the orientation of the ssDNA nucleotides, we used the EcSSB Thr98Cys mutant tagged with the TMS tag of Figure 1. The 1H-NMR signal of the TMS group was at about -0.28 ppm (Figure 4D). Its line width was 17 Hz. Following addition of dT67, the resonance shifted to -0.40 ppm (Figure 4E) and its width increased to 27 Hz. The free protein showed NOESY cross-peaks at similar chemical shifts as EcSSB Tyr97Tby, except that two new cross-peaks appeared at 1.6 and 1.8 ppm, which were more intense than all other intermolecular NOEs, indicating that contacts with methyl groups of dT67 are readily observable. Residue 98 is located more deeply in the binding pocket of the ssDNA binding groove of the OB domains than residue 97. These NOEs thus support the view that the ssDNA projects one or more pyrimidine rings into the binding groove at this site.

The t-buty1 group of a solvent-exposed O-t-buty1-tyrosine residue features rapid rotations of methyl groups and rotations around carbon-oxygen bonds, which permits the observation of a 1H-NMR signal even in a slowly tumbling protein. [8] The present work shows that a quite narrow 1H-NMR signal can also be obtained for a TMS group, which has a resonance frequency close to 0 ppm and thus little overlap with 1H resonances of the protein. The signals of t-buty1 and TMS groups can readily be identified, making further resonance assignments of the protein unnecessary. Either group can be deployed to observe site-specific intermolecular NOEs. Sensitivity of the NOESY experiment can be maximized by taking advantage of the narrow line shape of the signal and using relatively long acquisition times during the detection period, while short evolution times can be used in the indirect dimension to detect NOEs with the broad resonances of the interaction partners. The example of the 95 kDa EcSSB-dT67 complex demonstrates that such intermolecular NOEs can be detected in systems of high molecular weight without isotope labeling and with excellent sensitivity, although we used only dilute solutions containing high salt concentrations and 3 mm NMR tubes in a 5 mm probehead.

It is important to note that it is difficult to draw atomic level structural conclusions from the intermolecular NOEs in the presence of spin-diffusion, which is profound for high molecular weight systems at the mixing time used (200 ms). Furthermore, the chemical modifications may perturb native interactions. Finally, excessively broad 1H-NMR line widths may prevent the observation of NOEs. For example, signals of the H6 protons of dT67 may have been too broad to be observable. In general, however, broadening of intramolecular NOE cross-peaks in two dimensions actually facilitates the detection of intermolecular NOEs with t-buty1 and TMS groups, which are broad only in a single dimension. As even the signal of the TMS group may overlap with protein resonances, observation of a conserved set of NOE cross-peaks with two different chemical probes provides a good way to check for consistent results.

In conclusion, intermolecular NOEs with t-buty1 and TMS groups can be detected in systems of high molecular weight with remarkable ease without isotope labeling. This opens the door to gain site-specific information on macromolecular interactions, at least in a qualitative way, in situations where isotope labeling and resonance assignments may be costly or impossible, as is often the case, e.g., for proteins that can be produced only with eukaryotic expression systems.

Acknowledgements

We thank Professor Nicholas E. Dixon for discussions and an expression plasmid of wild-type SSB. Financial support by the Australian Research Council is gratefully acknowledged.

Keywords: nuclear magnetic resonance • nuclear Overhauser effect • single-stranded DNA binding protein • t-buty1 group • trimethylsilyl tag

COMMUNICATION


Making contact: Intermolecular nuclear Overhauser effects between large proteins and binding partners can be observed selectively and with high sensitivity following site-specific chemical tagging of the proteins with t-butyl or trimethylsilyl groups.