Biomolecular NMR Assignments

NMR backbone resonance assignment and solution secondary structure determination of human NSD1 and NSD2

--Manuscript Draft--

Manuscript Number:	
Full Title:	NMR backbone resonance assignment and solution secondary structure determination of human NSD1 and NSD2
Article Type:	Article (original research)
Keywords:	Nuclear receptor-binding SET domain; NSD1; NSD2; mmSET; WHSC1; NSD family; Histone-methyl transferase; NMR resonance assignments; NMR stability screen
Corresponding Author:	Glyn Williams, D.Phil. Astex Pharmaceuticals Cambridge, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Astex Pharmaceuticals
Corresponding Author's Secondary Institution:	
First Author:	Glyn Williams, D.Phil.
First Author Secondary Information:	
Order of Authors:	Glyn Williams, D.Phil.
	Nader Amin
	Daniel Nietlispach
	Seema Qamar
	Joe Coyle
	Elisabetta Chiarparin
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Proteins of the NSD family are histone-methyl transferases with critical functions in the regulation of chromatin structure and function. NSD1 and NSD2 are homologous proteins that function as epigenetic regulators of transcription through their abilities to catalyse histone methylation. Misregulation of NSD1 and NSD2 expression or mutations in their genes are linked to a number of human diseases such as Sotos syndrome, and cancers including acute myeloid leukemia, multiple myeloma, and lung cancer. The catalytic domain of both proteins contains a conserved SET domain which is involved in histone methylation. Here we report the backbone resonance assignments and secondary structure information of the catalytic domains of human NSD1 and NSD2.

NMR backbone resonance assignment and solution secondary structure determination of human NSD1 and NSD2

Nader Amin[1,4], Daniel Nietlispach[2], Seema Qamar[3], Joe Coyle[1], Elisabetta Chiarparin[1,5] and Glyn Williams[1 *]

[1] Astex Pharmaceuticals, 436 Cambridge Science Park, Cambridge CB4 0QA, UK

[2] Department of Biochemistry, University of Cambridge, 80 Tennis Court Rd, Cambridge CB2 1GA, UK

[3] Cambridge Institute for Medical Research, University of Cambridge, Hills Road, Cambridge CB2 0XY

[4] Present Address: Nuffield Department of Population Health, University of Oxford, Richard Doll Building, Richard Doll Building, Oxford, OX3 7LF

[5] Present Address: AstraZeneca, 310 Cambridge Science Park, Cambridge CB4 0WG, UK

Corresponding Author: *To whom correspondence should be addressed.

Glyn Williams, glyn.williams@astx.com

Abstract

Proteins of the NSD family are histone-methyl transferases with critical functions in the regulation of chromatin structure and function. NSD1 and NSD2 are homologous proteins that function as epigenetic regulators of transcription through their abilities to catalyse histone methylation. Misregulation of NSD1 and NSD2 expression or mutations in their genes are linked to a number of human diseases such as Sotos syndrome, and cancers including acute myeloid leukemia, multiple myeloma, and lung cancer. The catalytic domain of both proteins contains a conserved SET domain which is involved in histone methylation. Here we report the backbone resonance assignments and secondary structure information of the catalytic domains of human NSD1 and NSD2.

Keywords

Nuclear receptor-binding SET domain, NSD1, NSD2, mmSET, WHSC1, NSD family, Histone-

methyl transferase, NMR resonance assignments, NMR stability screen

Biological context

The nuclear SET domain (NSD) containing proteins are a family of mammalian histone-methyl transferases with diverse functions in cell growth and differentiation. This family comprises NSD1, NSD2 (also known as MMSET/WHSC1) and NSD3 (also known as WHSC1L1). These proteins are emerging as important therapeutic targets in the epigenetic therapy of cancers. Members of the NSD family, all of which contain a conserved Su(var) 3-9, enhancer of zeste, trithorax (SET) domain, act to methylate specific lysine residues on histone tails (Huang et al. 1998, Li et al. 2009, Rayasam 2003, Varier and Timmers 2010). This serves as an epigenetic marker for the recruitment of various complexes controlling chromatin organisation and gene expression. Both NSD1 and NSD2 regulate methylation of lysine 36 on histone 3, a process which is crucial to chromatin regulation. In addition, both NSD1 and NSD2 have been shown to regulate methylation of lysine 20 on histone 4 and the non-histone p65 subunit of nuclear factor-κB (NF-κB) (Berdasco et al. 2009, Lu et al. 2010, Rayasam et al. 2003, Wang et al 2007, Pei et al 2011). Aberrant expressions of NSD1 and NSD2 are associated with multiple types of cancer, including acute myeloid leukemia, multiple myeloma, and lung cancer (Jaju et al. 2001, Keats et al 2003, La Starza et al. 2004, Tonon et al. 2005, Wang et al. 2007). In addition to its role in cancer, mutations or alterations to the NSD1 gene are responsible for a number of developmental disorders, including Sotos and Weaver syndromes (Douglas et al. 2003).

Both NSD1 (2696 residues) and NSD2 (1365 residues) contain a number of functional domains, including a catalytic domain, two proline-tryptophan-tryptophan-proline (PWWP) domains, four zinc-finger plant homeodomains (PHDs), and two nuclear receptor interaction domains (Kurotaki et al. 2001, Qiao et al. 2011). The catalytic domain, which is responsible for histone methylation, is itself comprised of a SET domain and two cysteine-rich flanking domains; an N-teminal pre-SET

domain and a C-terminal post-SET domain. Although no tertiary structure yet exists for NSD2, the structure of the catalytic domain of NSD1 has been solved (Qiao et al. 2011). The SET domain is approximately 130 amino acids in length and contains adjacent binding sites for the lysine-[histone] ligand and the cofactor S-adenosylmethionine (SAM), which acts as a methyl group donor. The post-SET domain forms one side of the SAM binding pocket and is crucial for the catalytic competency of the enzyme. The loop connecting the SET and post-SET domains (the post-SET loop) is able to block the substrate-lysine binding site and regulate substrate binding.

The high sequence similarity between the catalytic domains of NSD1 and NSD2 (76% identity) indicates a similar overall topology, with pre-SET, SET and post-SET domains. Despite these structural insights on NSD1 and NSD2, the molecular bases of substrate specificity and catalysis are not fully understood. Here we report the backbone ^{1}H, ^{15}N and ^{13}C assignment of the catalytic domains of human NSD1 (residues 1852 – 2082) and NSD2 (residues 973 - 1203) and their secondary structure.

Methods and experiments

Protein expression and purification

The catalytic domains of human NSD1 (residues 1852 -2082) and NSD2 (residues 973-1203) were each produced with a TEV protease-cleavable His-tag at the N-terminus. These constructs were subcloned into pET23b vectors and overexpressed in E. *coli* Bl21 cells. Cleavage of the His-tag resulted in an additional three residues at the N-terminus (GSH). Uniformly ^{15}N-labelled and $^{15}N/^{13}C$-labelled protein were expressed in minimal medium supplemented with either $^{15}NH_4Cl$ (1 g/l) or $^{15}NH_4Cl$ (1 g/l) and ^{13}C-glucose (2 g/l) as the sole nitrogen and carbon sources, respectively. Unlabelled protein was expressed in Terrific Broth (TB). Cells were grown at 37 °C to an optical density at 600 nm of 0.5 and expression was induced with 1 mM IPTG. The cells were incubated at 15 °C for 16 hours. Cells were lysed by sonication and purified by immobilised metal affinity chromatography. The His-tag was removed using TEV protease and the proteins were further

purified by size-exclusion chromatography. The proteins were then concentrated and dialysed against the appropriate NMR buffer.

NMR spectroscopy:

The NMR data were acquired at either Astex Pharmaceuticals or the Department of Biochemistry, University of Cambridge. Each NMR sample was made up to a volume of 160 µl in a 2.5 mm capillary which was placed in a standard 5 mm NMR tube. At Astex Pharmaceuticals, experiments were carried out at 310 K unless otherwise stated, using a Bruker Avance III 500 MHz spectrometer equipped with a 5 mm TXI cryoprobe. NMR spectra were processed using TOPSPIN and analysed using either TOPSPIN or CCPN Analysis (Vranken et al. 2005). At the Department of Biochemistry, University of Cambridge, experiments were carried out at 310 K on a Bruker Avance III spectrometer equipped with a 5 mm QCI HFCN/z cryoprobe operating at a ^{1}H frequency of 600 MHz. The data was collected with non-uniform sampling (10% of the full matrix were sampled). The 3D spectra were reconstructed using the maximum entropy module implemented in the Azara processing software (W. Boucher, unpublished).

The folded state and stability of NSD1 and NSD2 catalytic domains were initially assessed using ^{1}H-^{15}N HSQC NMR spectra. The best quality spectra, as judged by line-widths and number of peaks, was obtained for both proteins at 310 K, with a saturating concentration of the co-factor SAM (2 mM). However, both proteins became significantly degraded at this temperature within a few hours. A number of buffer conditions were therefore tested for their effects on protein stability at 310 K. Protein stability was assessed by 1D proton NMR spectroscopy using unlabelled protein, by observing the attenuation of methyl and amide peaks in the 1D spectra over time. For the 1D ^{1}H NMR stability screen, all experiments were carried out using unlabelled protein at a concentration of 50 µM, with 2 mM SAM and a temperature of 310 K. The buffers used were either 50 mM MES or sodium phosphate, pH 6 or 6.5 and a range of salt (NaCl) concentrations between 150 and 450 µM.

The backbone ^{1}H, ^{13}C and ^{15}N resonances of NSD1 and NSD2 catalytic domains were assigned

using $^{13}C/^{15}N$- uniformly-labelled protein at a concentration of 200 μM. For NSD1 the NMR buffer used was MES 50 mM pH 6.5, NaCl 450 mM, TCEP 2 mM, SAM 2 mM, NaN_3 0.01% w/v, D_2O 10% v/v. For NSD2 the NMR buffer was sodium phosphate 50 mM pH 7, NaCl 450 mM, TCEP 2 mM, SAM 2 mM, NaN_3 0.01% w/v, D_2O 10% v/v. The protein was split into several aliquots which were stored at -20º C. Because of the poor stability of both proteins at 310 K, each sample was kept in the spectrometer at 310 K for a maximum of 2 days and was replaced with a fresh sample if necessary. The experiments were recorded at the Department of Biochemistry, University of Cambridge and were a combination of standard triple resonance experiments; a 2D $^{1}H-^{15}N$ HSQC, and 3D HNCO, HNCA, HN(CO)CA, HN(CO)CACB and HNCACB experiments (Sattler et al. 1999). Except for the HNCO all experiments were repeated several times (2-3 times) using fresh samples and the spectra co-added to improve the overall sensitivity.

Methods and experiments

Extent of assignments and data deposition

Figures 1 and 2 show the $^{1}H-^{15}N$ HSQC spectra at 310 K of the catalytic domains of human NSD1 and NSD2, respectively. The spectra indicate that both proteins are well-folded. Each construct contains an additional four residues at the N-terminus, which were part of the hexahistidine tag that remained on the protein following cleavage of the tag. These residues were not assigned and are therefore not identified in either HSQC spectrum. For NSD1, 212 of the 219 non-proline amide residues (96.8%) of the catalytic domain were assigned. 97.8% of the Cα resonances, 80.5% of the Cβ resonances and 90.4 % of the carbonyl C resonances were assigned. For NSD2, 206 of the 219 non-proline amide residues of the catalytic domain were assigned (94.1 %). 95.7 % of the Cα resonances, 93.5% of the Cβ resonances and 88.8 % of the carbonyl C resonances were assigned.

The chemical shifts have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession numbers 26767 and 26768 for NSD1 and NSD2 respectively. For NSD1, seven residues

of the core construct could not be assigned (Q1858, G1929, L1950, Q1951, C2022, C2023 and Q2024). Most of these resonances appear to be broadened beyond detection in the 3D experiments, suggesting conformational exchange on an intermediate timescale in these regions. With the exception of Q1858 and G1929, these residues are situated on loops in the SET domain adjacent to the SAM binding site. The deviations in backbone $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts from random coil values are shown in figure 3. These are broadly consistent with the secondary structure of the crystalline protein (Qiao et al. 2011).

Secondary structure determination

The two protein sequences are highly similar. A pairwise sequence alignment using the Needleman and Wunsch algorithm (Needleman and Wunsch, 1970) (figure 3) yielded 76.3% identity and 89.2% similarity over 232 residues. Indeed, the similarity was such that no gaps were introduced into the alignment. Figure 3 shows the secondary structures of both proteins, derived from their backbone chemical shifts. Also shown in this figure is the secondary structure of human NSD1, as determined from its crystal structure (PDB ID 300I) (Qiao et al. 2011). The NMR-derived secondary structures of both proteins correspond well with the expected fold for the catalytic domain, with a core SET domain motif composed of three beta sheets with an alpha-helix adjacent to the second beta sheet. This core SET domain is flanked by an N-terminal helix adjacent to beta sheet 1 and the AWS domain at the other end, adjacent to beta-sheet 3.

There is broad general agreement between the NMR-derived secondary structures of the two proteins, with the exception of the N-terminus, which is less well defined in NSD2. The NMR-derived secondary structures agree well with the secondary structure of NSD1 derived from the crystal structure. The two proteins are therefore expected to have a similar topology and fold, with no major structural differences between them. It is hence postulated that difficulties encountered during crystallisation of NSD2 are due to increased conformational flexibility of this protein

compared to NSD1.

References

Berdasco M, Ropero S, Setien F, Fraga MF, Lapunzina P, Losson R, Alaminos M, Cheung N, Rahman N, Esteller M (2009) Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 inhuman neuroblastoma and glioma. Proc. Natl. Acad. Sci. U. S. A. 106(51): 21830-21835.

Douglas J, Hanks S, Temple IK, Davies S, Murray A, Upadhyaya M, Tomkins S, Hughes HE, Cole TR, Rahman N (2003) NSD1 mutations are the major cause of Sotos syndrome and occur in some cases of Weaver syndrome but are rare in other overgrowth phenotypes. Am. J. Hum. Genet. (72) 132–143.

Huang N, Vom Baur E, Garnier JM, Lerouge T, Vonesch JL, Lutz Y, Chambon P, Losson R (1998) Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators, EMBO J. (17) 3398–3412.

Jaju RJ, Fidler C, Haas OA, Strickson AJ, Watkins F, Clark K, Cross NC, Cheng JF, Aplan PD, Kearney L, Boultwood J, Wainscoat JS (2001) A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. Blood (98) 1264–1267.

Keats JJ, Reiman T, Maxwell CA, Taylor BJ, Larratt LM, Mant MJ, Belch AR, Pilarski LM (2003) In multiple myeloma, t(4;14)(p16;q32) is an adverse prognostic factor irrespective of FGFR3 expression. Blood. (101) 1520–1529.

Kurotaki N, Harada N, Yoshiura K, Sugano S, Niikawa N, Matsumoto. (2001) Molecular characterization of NSD1, a human homologue of the mouse Nsd1 gene. N. Gene (279) 2:197-204.

La Starza R, Gorello P, Rosati R, Riezzo A, Veronese A, Ferrazzi E, Martelli MF, Negrini M, Mecucci C (2004) Cryptic insertion producing two NUP98/NSD1 chimeric transcripts in adult refractory anemia with an excess of blasts. Genes Chromosom Cancer (41) 4: 395-399.

Li Y, Trojer P, Xu CF, Cheung P, Kuo A, Drury III WJ, Qiao Q, Neubert TA, Xu RM, Gozani O, Reinberg D (2009) The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. J. Biol. Chem. (284) 34283–34295.

Lu T, Jackson MW, Wang B, Yang M, Chance MR, Miyagi M, Gudkov AV, Stark GR (2010) Regulation of NF-kappaB by NSD1/FBXL11-dependent reversible lysine methylation of p65, Proc. Natl. Acad. Sci. U. S. A. (107) 46–51.

Needleman SB, Wunsch CD (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48 (3):443–53.

Pei H, Zhang L, Luo K, Qin Y, Chesi M, Fei F, Bergsagel PL, Wang L, You Z, Lou Z (2011) MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. Nature, 470 (7332) : 124 128).

Qiao (Q), Li Y, Chen Z, Wang M, Reinberg D, Xu RM (2010) The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation, J. Biol. Chem. (286) 10:8361-8368.

Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, Lerouge T, Hager GL, Chambon P, Losson R (2003) NSD1 is essential for early post- implantation development and has a catalytically active SET domain, Embo J. (22) 3153–3163.

Sattler M, Schleucher, J, Griesinger, C (1999) Heteronuclear multidimensional NMR methods for the structure determination of proteins in solution employing pulsed field gradients. Prog NMR Spectrosc 34:93-158.

Varier RA, Timmers HT (2011) Histone lysine methylation and demethylation pathways in cancer Biochim. Biophys. Acta. (1815) 1:75-89.

Wang GG, Cai L, Pasillas MP, Kamps MP (2007) NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis, Nat. Cell Biol. (9) 804–812.

Figure

Click here to download Figure MMSET NSD1 JBNMR Figures 24Mar2016.docx

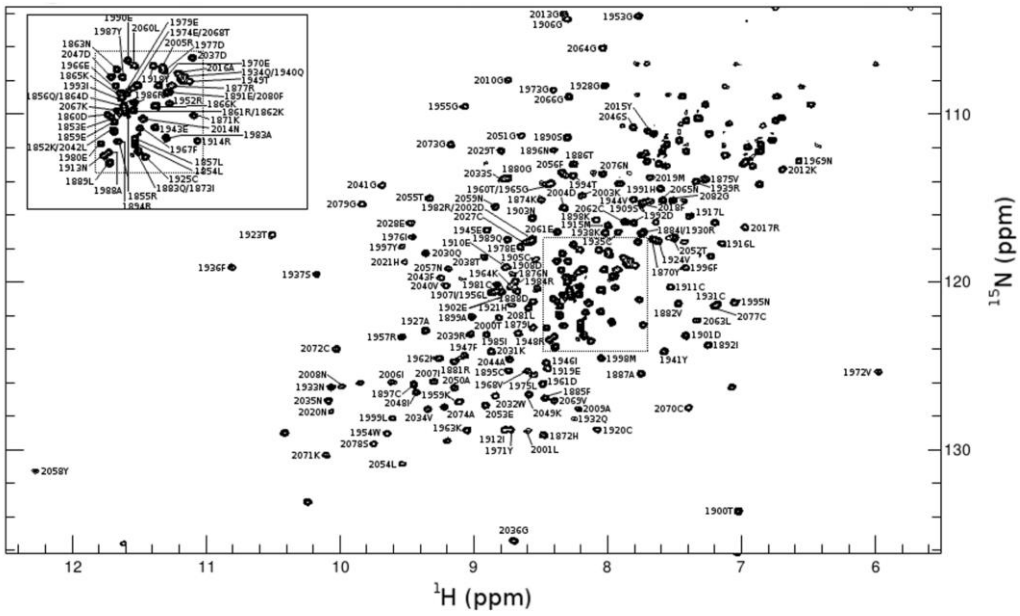

Figure 1. ^{1}H-^{15}N HSQC spectrum of human NSD1 catalytic domain. The spectrum was recorded using 200 μM ^{13}C-^{15}N-labelled protein at 600 MHz. Residue assignments are indicated. For the sake of clarity, a number of peaks are shown annotated separately.

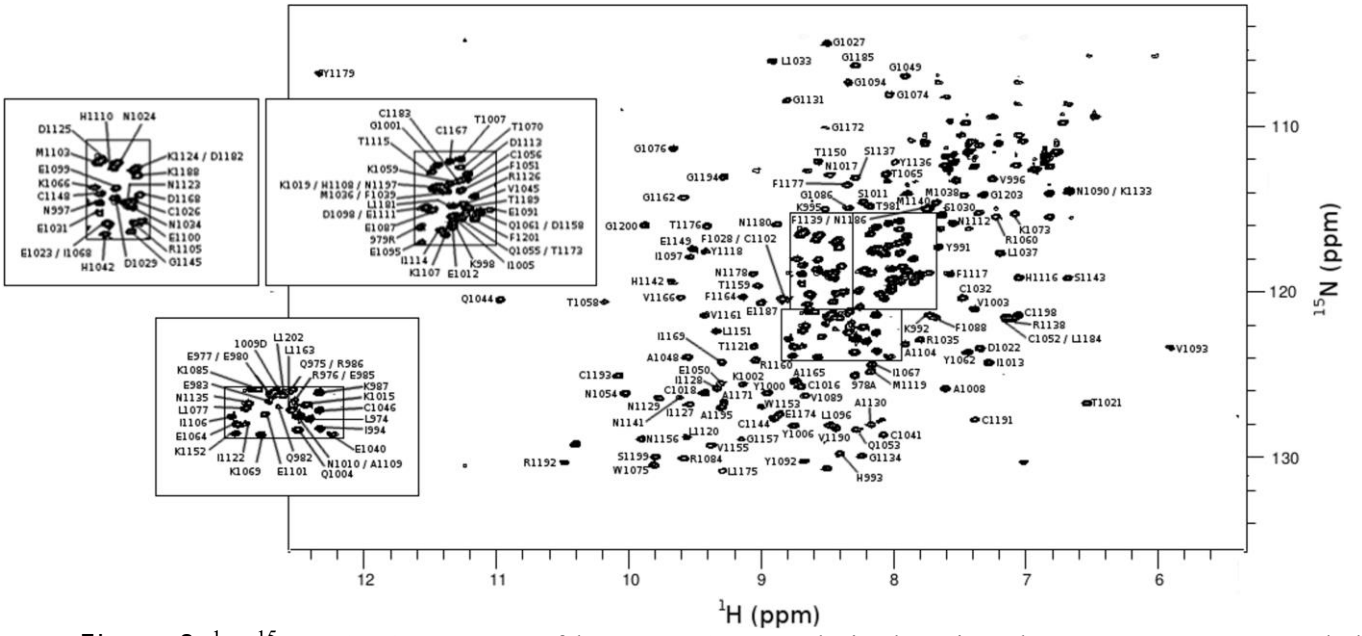

Figure 2. ^{1}H-^{15}N HSQC spectrum of human NSD2 catalytic domain. The spectrum was recorded using 200 μM ^{13}C-^{15}N-labelled protein at 600 MHz. Residue assignments are indicated. For the sake of clarity, a number of peaks are shown annotated separately

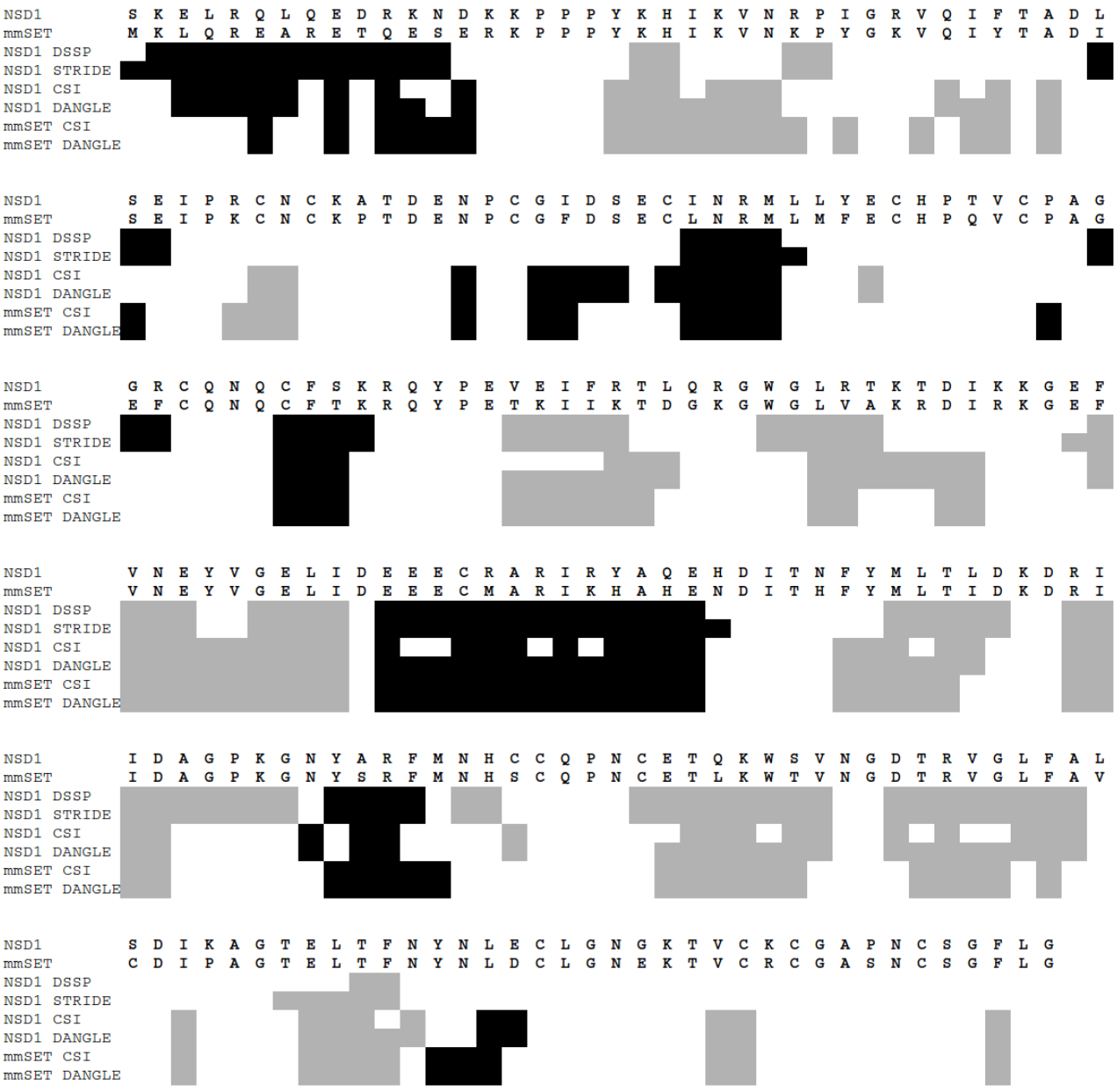

Figure 3. Sequence alignment between NSD1 and NSD2, illustrating similarities in secondary structure. Alpha-helices and beta-sheets are depicted as black and grey squares, respectively. In each block: Rows 1 and 2 show the sequences of NSD1 and NSD2 respectively. Rows 3 and 4 show the secondary structure of NSD1 as derived from the crystal structure. These have been defined using the programs DSSP (Kabsch and Sander, 1983) (row 3) and STRIDE (Heinig and Frishman, 2004) (row 4). Rows 5 and 6 contain the NMR-derived secondary structure of NSD1, calculated using the programs CSI (Hafsa and Wishart 2014) and DANGLE (*Cheung, Maguire et. al., 2010*), respectively. Rows 7 and 8 contain the NMR-derived secondary structure of NSD2, calculated using the programs CSI and DANGLE, respectively.