From NSD1 to Sotos syndrome : a genetic and functional analysis

Visser, R.

Chapter 5

Non-hotspot-related breakpoints of common deletions in Sotos syndrome are located within destabilized DNA regions

Remco Visser[1,2,3,4,5], Osamu Shimokawa[1,4,6], Naoki Harada[1,4,6], Norio Niikawa[1,4] and Naomichi Matsumoto[3,4]

J Med Genet 2005; 42: e66

1. Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan
2. Department of Pediatrics, Leiden University Medical Center, Leiden, The Netherlands
3. Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan
4. CREST, Japan Science and Technology Agency, Kawaguchi, Japan
5. International Consortium for Medical Care of Hibakusha and Radiation Life Science, The 21st Century COE (Center of Excellence), Nagasaki, Japan
6. Kyushu Medical Science Nagasaki Laboratory, Nagasaki, Japan

Abstract

Background

Sotos syndrome (SoS; OMIM 117550) is an disorder characterized by excessive growth, typical craniofacial features and developmental retardation. It is caused by haploinsuffiency of *NSD1* at 5q35. There is a 3.0-kb recombination hotspot in which the breakpoints of around 80% of SoS patients with a common deletion can be mapped.

Objective

To identify deletion breakpoints located outside the SoS recombination hotspot.

Methods

A screening system for the directly orientated segments of the SoS low-copy repeats was developed for 10 SoS patients with a common deletion, but who were negative for the SoS hotspot. Deletion-junction fragments were analyzed for their DNA duplex stability and their relation to scaffold/matrix attachment regions (S/MARs). These features were compared with the SoS hotspot and recombination hotspots of other genomic disorders.

Results

The breakpoint in four SoS patients, two with a deletion in the maternally derived chromosome. These breakpoint regions were located ~2.5 kb, ~9.6 kb, ~27.2 and ~27.7 kb telomeric to the SoS hotspot and were confined to 164 bp, 46 bp, 256 bp and 124 bp, respectively. Two of the regions were mapped within *Alu*-elements. All crossover events were found to have occurred within or adjacent to a highly destabilized DNA duplex with a high S/MAR probability. In contrast, the SoS hotspot and other genomic disorders' recombination hotspots were mapped to stabilized DNA helix regions, flanked by destabilized regions with high probability to contain S/MAR elements.

Conclusion

These data suggest that a specific chromatin structure might increase susceptibility for recurrent crossover events and consequently predispose for recombination hotspots in genomic disorders.

Introduction

Sotos syndrome (SoS; OMIM 117550) is a congenital disorder characterized by overgrowth, distinctive craniofacial features and various degrees of developmental delay (1). Aberrations of the *N*uclear receptor binding *S*ET *D*omain protein *1* (*NSD1*) at 5q35 include intragenic mutations or submicroscopic whole-gene deletions (2-9). Microdeletions are found in around 50% of the Japanese SoS patients population, while they account for about 10% of the non-Japanese SoS patients (10). Recently, we showed that the 1.9-Mb common microdeletion is caused by homologous recombination between directly orientated segments (PLCR-B and DLCR-2B) of the proximal and distal low-copy repeats (PLCR and DLCR) (11). The unequal strand exchange region was limited to a 3.0-kb hotspot in which we mapped the breakpoints of 78.7% (37/47) of our Sotos patients with a common deletion. This major hotspot was recently confirmed by others (12). Similar analysis at a nucleotide level of recombination hotspots in other genomic disorders has identified, amongst others, regions of uninterrupted sequence homology, several sequence motifs and raised GC content as hotspot features (11,13-16). However, these features are not consistent for all hotspots and, owing to the analytically difficult background of highly homologous LCRs, the number of identified hotspot- and non-hotspot-related breakpoints is limited. Other possible contributing factors, such as epigenetic alterations or specific chromatin structure, have been suggested (17). Interestingly, breakpoints of gross deletions were indeed found to coincide with non-B DNA conformations (18). Non-B DNA conformation could result in an increase of accessibility for cleavage enzymes or a weakened chemical stability of the DNA helix, or both (18). Recently, breakpoints of recurrent intragenic deletions of the *R*etinoblastoma *1* (*RB1*) gene were located within a transition region between double-stranded B-DNA and single-stranded DNA (19). This region was adjacent to a strong scaffold/matrix attachment region (S/MAR) (19). S/MARs are responsible for chromatin attachment to the nuclear matrix and for organisation of chromatin into loop domains (20). Thus, chromatin organisation in relation to stability of the DNA duplex might be a contributing factor for hotspot predisposition in genomic disorders.

In this study, we screened the directly orientated regions within the Sotos LCRs in order to identify deletion breakpoints located outside the SoS recombination hotspot. The deletion-junction fragments found were investigated at a nucleotide level and compared with the SoS hotspot in regard to their locations, neighboring structures, stability of the DNA helix

(so called stress-induced destabilization duplex (SIDD)) and probability of containing an S/MAR element. Furthermore, the recombination hotspots of other genomic disorders were analyzed for their SIDD and S/MAR profiles.

Material and methods

Patients

This study included 10 Japanese patients with SoS who carry a common deletion but from whom the breakpoint could not be mapped to the SoS hotspot (3,11). Furthermore, available parental DNA of patients with newly identified breakpoints was analyzed. The control group consisted of 50 healthy Japanese individuals. After informed consent, genomic DNA was obtained from peripheral blood cells or lymphoblastoid cell lines using standard methods. Experimental protocols were approved by the Committee for Ethical Issues at Yokohama City University School of Medicine and by the Committee for Ethical Issues on Human Genome and Gene Analysis at Nagasaki University.

Screening by long-range polymerase chain reaction

Methods followed were similar to those as described previously (11). In short, sets of primers with the forward primer specific for PLCR-B and the reverse for DLCR-2B were designed with the online version of Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (21). Primer sequences and product length are shown in Table 1.

Amplification was tested on PLCR-B BAC-clone RP11-546L14 (GenBank accession number AC108509), DLCR-2B BAC-clone CTD-2515I1 (GenBank accession number AC118457) and genomic DNA from a normal individual. The annealing temperatures decisive for specific amplification of a possible deletion-junction fragment were determined experimentally. Long-range polymerase chain reaction (PCR) was performed using the GeneAmp XL PCR Kit (Applied Biosystems, Foster City, California, USA). Positive PCR products were amplified with nested primers and subsequently sequenced. For primer set 6, all products were first submitted to restriction with *Fsp*I to eliminate the amplified product of the normal DLCR-2B and possible breakpoint-junction fragments were sequenced. All nested primer sequences and conditions are available upon request. Paralogous sequence variants (PSVs) (i.e., nucleotide difference between the PLCR-B and DLCR-2B) (22) were mapped to the

Table 1. Primer sequences used for long-range PCR reactions

Set	Forward primer	Reverse primer	Product length (kb)
1	CCAGCGTTATATGTTCAGTCCTAGATGAAG	GCAAAACTGCCGTCCCTCAC	8.3
2[1]	TGGTCTGATTCCTATGTTCTGCTGGtTG[2]	CCCAGTGCTGGGGCACAAGTgA[2]	6.8
3[1]	CACCAAAGGCCAGTGATGCCAATA	AGCCCTCCCCTGGCCGACTG	6.9
4	GGCCAGTGCATGATGTAGTCA	CAGTCACTGATGCTAACCTTGGAT	11.1
5	TTCCTCAGAGAGGCTTCGTTTGtA[2]	GCTGGGTCCACCTGCATCTA	10.5
6	CCCATGTTCAAAGCACAACAgA[2]	TCCACCCCCAGGAAACAGAT	15.8

[1] Primer sets 2 and 3 were reported previously (11).

[2] Mismatched nucleotides introduced to increase specificity of the long-range PCR, are shown in small characters

PLCR-B and DLCR-2B according to the NCBI build 35 (May 2004) database (http://genome.ucsc.edu/).

Analysis of the deletion-junction fragments and recombination hotspots

The identified SoS deletion-junction fragments were analyzed including 3.0-kb of their flanking sequences. Repetitive sequence elements were identified with RepeatMasker (http://www.repeatmasker.org/). The sequences covering the recombination hotspots of deletions in neurofibromatosis type 1 (NF1; OMIM 162200) (14), of deletions in hereditary neuropathy with liability to pressure palsies (HNPP; OMIM 162500) and its reciprocal duplications in Charcot-Marie-Tooth disease type 1A (CMT1A; OMIM 118220) (23-25), of deletions in Smith-Magenis syndrome (SMS; OMIM 182290) and its reciprocal duplications in dup(17)(p11.2p11.2) (15,26) were obtained by use of the 'PCR' and 'Blat' functions on the UCSC homepage, containing the NCBI build 35 (May 2004 version). WebSIDD was used for the prediction of stress-induced-duplex-destabilized (SIDD) sites in double-stranded DNA (http://orange.genomecenter.ucdavis.edu/benham/sidd/) (27). Scaffold/Matrix attachment regions were predicted with S/MAR-Wiz version 1.0 (http://www.futuresoft.org/MAR-Wiz/) (28). Both programs were run under default conditions.

Results

Four primer sets were designed (Table 1) and in combination with the previously designed hotspot primer sets (11), a nearly complete coverage was achieved of PLCR-B and PLCR-2B (Figure 1A and 1B). Remaining small gaps included ~1.5-kb, ~0.4 kb and ~0.6 kb, respectively, due to difficulties in obtaining amplification of these regions. SoS 85 and SoS 110 showed a ~11.1 kb amplified product for primer set 4, while their respective parents were negative for the same reaction (Figure 2A). This indicated a deletion-junction fragment. Also 50 control samples were negative for this product. Sequencing revealed a transition of PSVs mapped to PLCR-B and those mapped to DLCR-2B for both patients. The breakpoint region for SoS 85 could be restricted to 164 bp (between nucleotide position 1319 and 1483) and to 46 bp for SoS 110 (between nucleotide position 8517 and 8563) (Figure 2C). PSVs at position 8975 and 9460 for SoS 110 are likely to be polymorphisms as they were also mapped to PLCR-B in SoS 85 (data not shown). The breakpoints were located ~2.5 kb and ~9.6 kb in the telomeric region of the SoS recombination hotspot for SoS 85 and SoS 110, respectively. SoS 4 and SoS 5 were after restriction with *Fsp*I positive for an ~11.3 kb product, which indicated a breakpoint-junction fragment. Restriction with *Fsp*I was necessary since also amplification of the normal DLCR-2B chromosome was occasionally detected. Fifty normal controls were screened and in seven an amplified product could be obtained. However, none of the seven controls showed a breakpoint-junction fragment after restriction with *Fsp*I. Unfortunately, parental DNA could not be obtained. The breakpoint-junction fragments were sequenced and the crossover regions based on PSVs were confined to 256 bp for SoS 5 (between position 5505 and 5761) and 124 bp for SoS 4 (between position 6028 and 6152) (Figure 2D). The two breakpoint regions are located ~27.2 kb and ~27.7 kb telomeric to the SoS hotspot and are separated from each other by ~0.3 kb. In SoS 5, an insertion of 4 nucleotides (GACA) was found at position 5594. This could indicate either the exact breakpoint location or a mere polymorphism.

For SoS 5 and SoS 85, the breakpoint region was mapped to a sequence not related to any interspersed repeats. However in SoS 85, a simple repeat (TA)n was found at position 1308-1319 and a LINE1-element was found in close proximity, that is 5 bp telomeric. The crossover event for SoS 4 occurred within an *Alu*-Sx element and the region for SoS 110 overlapped partially with an *Alu*-Sg element. In the breakpoint regions for the four patients, only in SoS 5 one translin motif (5'- GCCCWSSW-3') was detected. This motif was found increased for the

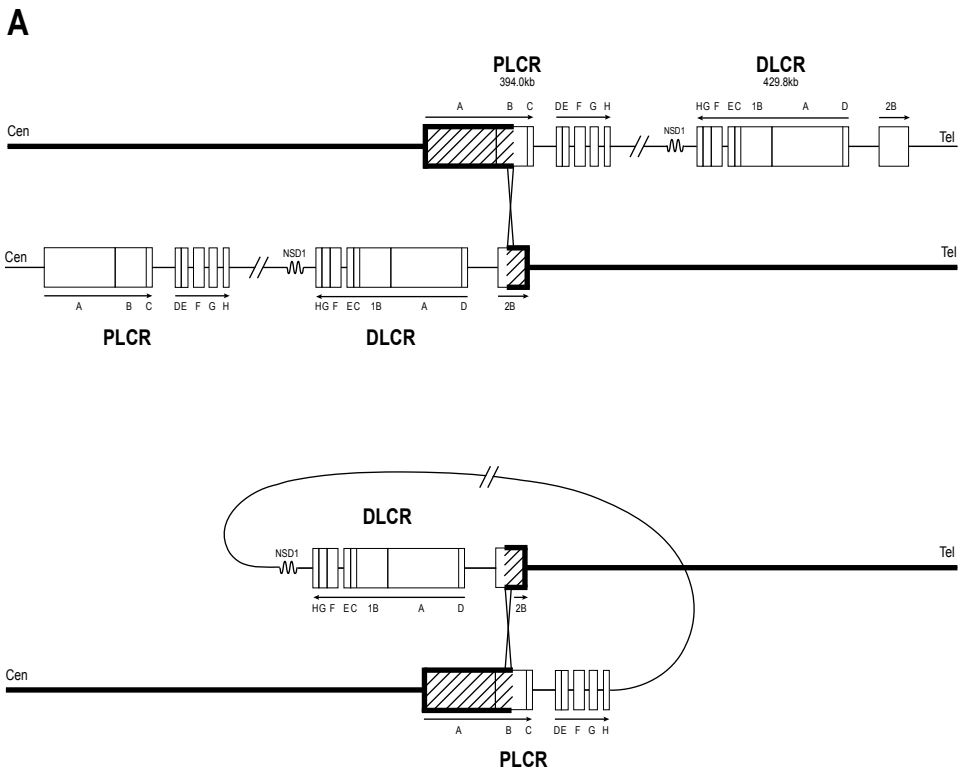


Figure 1A. Schematic presentation of possible non-allelic homologous recombination (NAHR) events resulting in the common 1.9-Mb microdeletion in SoS

Mechanisms of genomic rearrangements are reviewed in detail by Stankiewicz and Lupski (40). The upper part depicts the possible crossover in an interchromosomal or in an intrachromosomal recombination event. The lower part shows an intrachromatid crossover event. The predicted deletion-junction fragment is shown with thick black lines. The segments within the PLCR together with corresponding homologous counterparts in the DLCR, are indicated with blocks and their respective letters. PLCR-B is represented twice in the DLCR (DLCR-1B and DLCR-2B). The arrows indicate the genomic orientation.

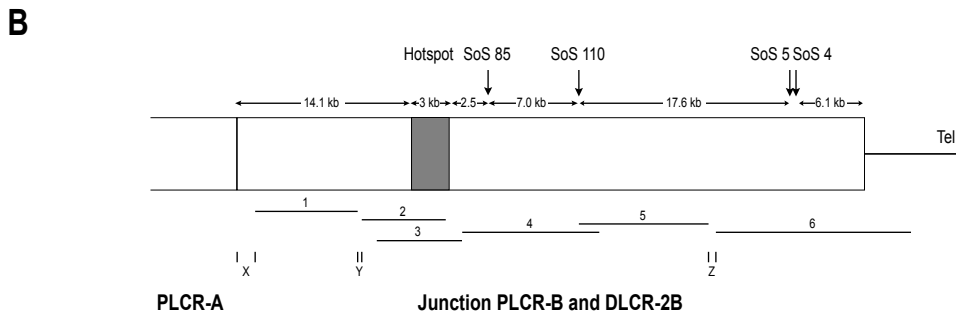


Figure 1B. Presentation of the deletion-junction fragment of directly orientated PLCR-B and DLCR-2B
A shaded box indicates the SoS hotspot. Vertical arrows show the breakpoint location of SoS 85, SoS 110, SoS 5 and SoS 4. Bidirectional arrows above the fragment depict the genomic distances between the different breakpoint locations and the remaining parts. Horizontal black lines below the deletion-junction fragment show the schematic position and length of long PCR products with the used primer sets. The letters X, Y, Z indicate existing gaps of ~1.5 kb, ~0.4 kb and ~0.6 kb, respectively. Cen: centromere; Tel: telomere.

SoS hotspot (11). Patients SoS 85 and SoS 110 were haplotyped previously and confirmed to carry a deletion in the maternally derived chromosome (29). The parental origin was unknown in SoS 4 and SoS 5. The deletion-junction fragment of SoS 85 arose through an intrachromosomal recombination event (Figure 1A lower), while this is not known for SoS 110 (29). The parents of SoS 85 and SoS 110 had a heterozygous inversion of the interval between the SoS LCRs (11). The father of SoS 5 showed also a heterozygous inversion, while the mother didn't carry an inversion (11). Parental DNA could not be obtained for SoS 4.

The results of the SIDD and S/MAR analysis based upon the proximal sequences of the involved PLCRs are shown in Figure 3. Analysis of the distal sequences of the respective LCRs did not show any significant differences in SIDD and S/MAR profiles owing to the high homology of proximal and distal LCRs (data not shown). The breakpoints of SoS 85 and SoS 110 (Figure 3A and B, respectively) overlapped with DNA regions which are very susceptible for duplex destabilization with G(x) values (i.e. the needed energy to force the base pair at position x always to be unpaired) (27) close to 0 kcal/mol. In concordance, the same regions showed increased S/MAR potential. The breakpoints of SoS 5 and SoS 4 were mapped to a transition region with an increased S/MAR potential and directly adjacent to destabilized DNA (Figure 3C). The SoS hotspot region was mapped to a ~4.8 kb segment of

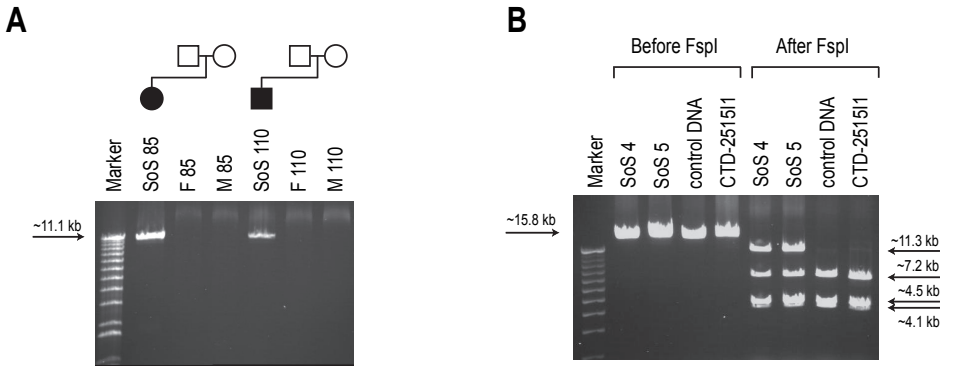


Figure 2A. PCR results for primer set 4 in patients SoS 85 and SoS 110 and their parents
Left lane: a 1-kb plus DNA ladder (Invitrogen, San Diego, California, USA). F 85: father of SoS 85; M 85: mother of SoS 85; F 110: father of SoS 110; M 110: mother of SoS 110.

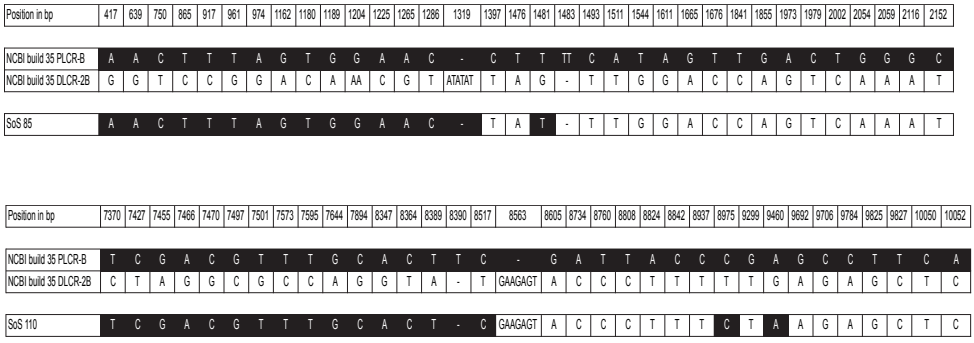
Figure 2B. PCR results for primer set 6 before and after restriction with *Fsp*I
Samples shown are SoS 4, SoS 5, DNA of a normal individual and clone CTD-2515I1. The detected ~11.3 kb product in SoS 4 and SoS 5 indicates a breakpoint-junction fragment. Left lane: a 1-kb plus DNA ladder.

highly stabilized DNA without S/MAR potential (Figure 3D). Also the recombination hotspots for NF1 (2.1 kb) (14), for common 4-Mb deletions in SMS (~8 kb) (15), and for uncommon deletions in SMS (577 bp) (26) showed similar stretches of stabilized SIDD sites covering the hotspots, flanked by non-stable regions with a high S/MAR potential (Figure 3E, F and G, respectively). The recombination hotspot in CMT1A and HNPP (557-741bp) (23-25) was mapped to a stabilized region, although this region also showed a slightly increased S/MAR potential (~0.10) (Figure 3H). However, the S/MAR potential of the corresponding sequence of the distal LCR was close to zero (data not shown).

Discussion

Common deletions in SoS syndrome are caused by non-allelic homologous recombination (NAHR) between directly orientated LCR segments (11). By use of long-range PCR screening, we identified the first non-hotspot-related breakpoints in four SoS patients with a common microdeletion. These breakpoint locations are expected to have a low recurrent character.

C



D

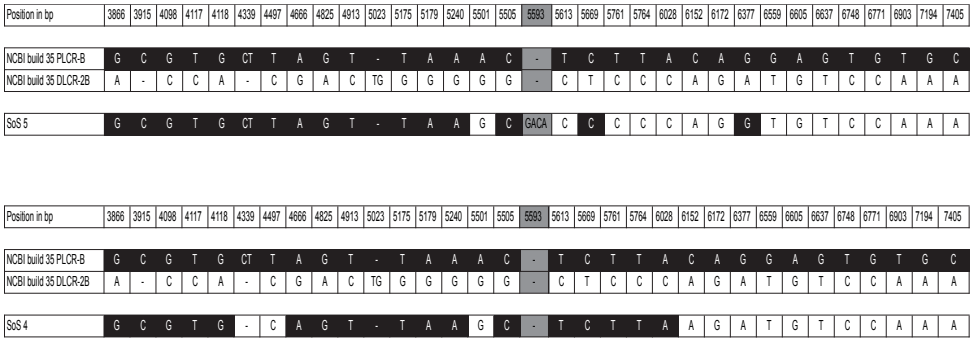


Figure 2C. PSVs identified in the breakpoint regions of SoS 85 (upper) and SoS 110 (lower)
Black boxes indicate PSVs of PLCR-B and white boxes show those of DLCR-2B. The PSVs as deposited in the NCBI build 35 (May version 2004) are shown above the respective patient's PSVs. The position in base pairs (bp) indicates the position of the PSVs within the product amplified with primer set 4.

Figure 2D. PSVs identified in the breakpoint regions of SoS 5 (upper) and SoS 4 (lower)
The position in base pairs indicates (bp) the position of the PSVs within the product amplified with primer set 6 and after restriction with *Fsp*I. The gray boxes show the position of the 4 inserted nucleotides as found in SoS 5.

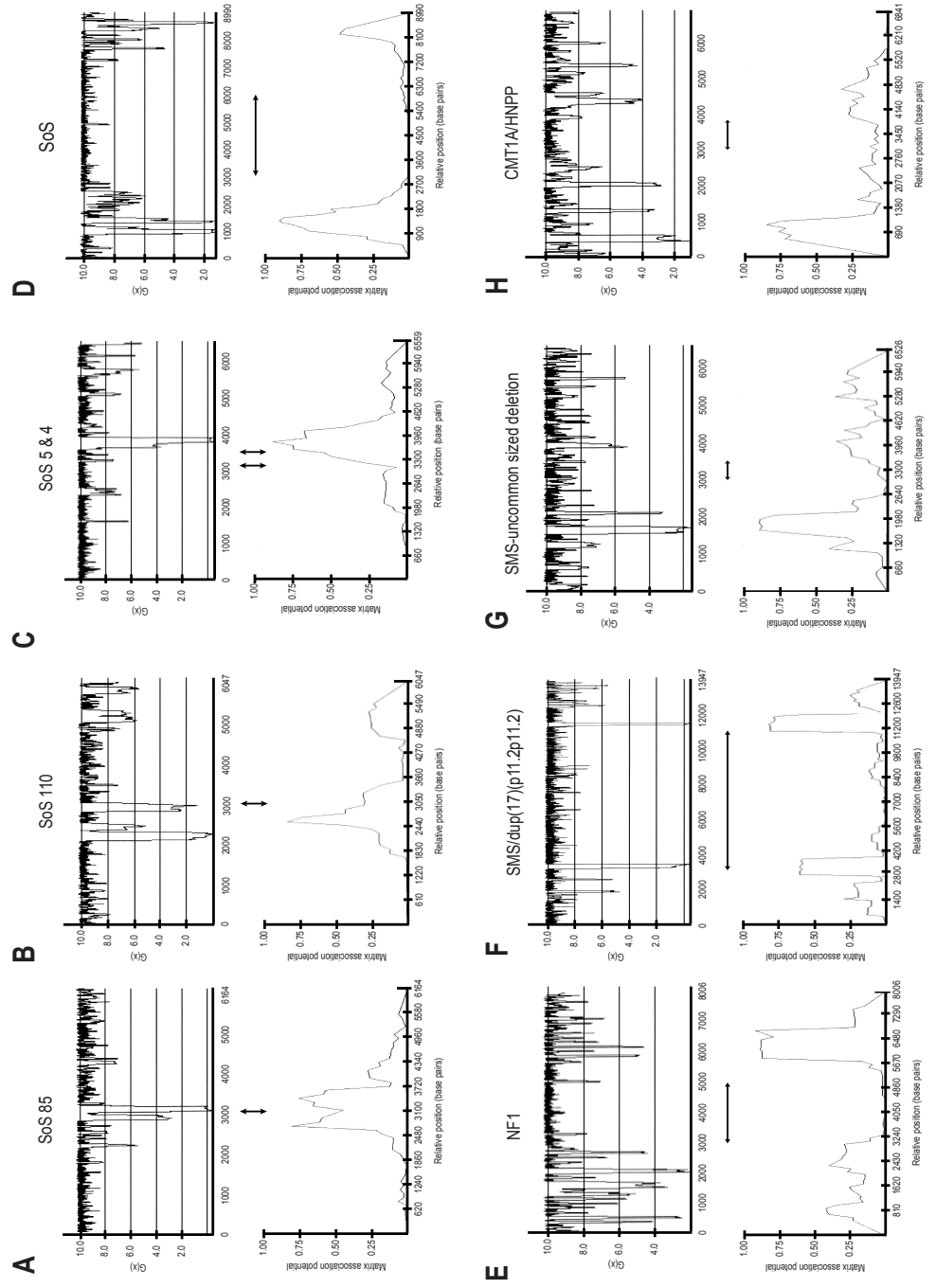

A
SoS 85
B
SoS 110
C
SoS 5 & 4
D
SoS
E
NF1
F
SMS/dup(17)(p11.2p11.2)
G
SMS-uncommon sized deletion
H
CMT1A/HNPP
G(x)
Matrix association potential
Relative position (base pairs)

Figure 3. SIDD destabilization energy profiles and S/MAR potentials

The SIDD destabilization energy profile (upper graph) and S/MAR potential (lower) of the proximal sequences of the involved LCRs are shown in relation to the breakpoint region in SoS 85 (A), SoS 110 (B), SoS 4 (C) and SoS 5 (C), (which are ordered based upon their genomic location in DLCR-2B (see Figure 1B)], the SoS hotspot (D), NF1-hotspot (E), SMS hotspot (F), SMS hotspot for uncommon sized deletions (G) and the CMT1A/HNPP hotspot (H). Analysis of the distal LCR sequences showed similar patterns (data not shown). Horizontal bidirectional arrows between the SIDD and S/MAR profiles indicate the respective hotspots. Vertical bidirectional arrows show the breakpoint regions for SoS 85 SoS 110, SoS 5 and SoS 4, respectively. In the SIDD profile, the y-axis shows the incremental free energy G(x) in (kcal/mol) which corresponds to the necessary energy to force the base pair at position x always to be open (27). For example, a G(x) value of 10.2 kcals/mol indicates unstressed, stabile DNA at that position. The y-axis in the S/MAR profile indicates the normalized S/MAR association potential with values between 0 and 1.0. The higher the predicted value, the more likely it is that the corresponding region contains a S/MAR element. Both x-axes show the sequence distances in base pairs.

Firstly because of the low frequency in our SoS patients population so far (2.1% (1/47)) for each of the breakpoint regions of SoS 85 and SoS 110 and 4.3% (2/47) for the region containing the breakpoints of SoS 4 and SoS 5. Secondly in case of SoS 85 and SoS 110, because of a maternal deletion origin, while all other patients previously haplotyped carried the deletion in the paternally derived chromosome (29). A rare maternal deletion origin and uncommon breakpoint locations are suggestive for a sex-biased recombination mechanism in these two patients, but investigations in larger populations are necessary for confirmation. It is however this low-recurrence characteristic in all four patients in combination with NAHR as the general underlying mechanism, which makes them interesting candidates for comparison with recombination hotspot features.

In SoS 5 and SoS 85, similar to the hotspot, the breakpoints were not located within short repetitive DNA elements. However, a LINE1 element was found in close proximity in SoS 85. Active human LINE1 retrotransposons *in vivo* have been shown to induce genomic instability such as inversions, deletions, and recombination between L1 elements (30,31). Although, because of truncating mutations (data not shown) within the two ORFs necessary for retrotransposition, it is unlikely that the LINE1 element near the breakpoint of SoS 85 maintained its function. On the other hand, the location of LINE elements has also been proposed to be a marker for the localization of S/MARs, as easily unwinding DNA might predispose for their integration within the genome (20). In SoS 4, the crossover event

occurred in an *Alu*-Sx element and in SoS 110 the breakpoint region partially overlapped with an *Alu*-Sg element. *Alu*-mediated illegitimate recombination is estimated to be responsible for ~0.3% of all human genetic diseases (32). An *Alu*-Sq/x element was found in the hotspot of uncommon sized deletions in SMS and the recombination interval was mapped to this element in 4 patients (26). However, this region was also characterized by a stabilized DNA helix and low S/MAR potential (Figure 3G).

The breakpoints of the other 6 SoS patients could not be identified. Although our LCR-specific long-range PCR proved its use, it is possible that the remaining primer sets are not sensitive enough, while only detection of a true positive can confirm their sensitivity. Also possible polymorphisms at the primer sites, unknown complex rearrangements within the LCRs during deletion-junction formation or a location of the breakpoints within the remaining gaps, might have inhibited detection.

To date, analysis of hotspot- and non-hotspot-related breakpoints have deepened our knowledge about the underlying causative mechanisms. LCRs in direct orientation with a high sequence identity are the necessary structures for rearrangement resulting in deletion and reciprocal duplication (33). Even higher sequence similarity is usually found within hotspots in combination with regions of uninterrupted sequence homology (~300-500 bp), which are proposed to be necessary for efficient recombination in mammalian cells (34,35). Several sequence motifs have been described, but neither has a common recombination initiating factor been found, nor have the identified motifs been confirmed *in vivo* (11,14,15,24,26). In the quest for an explanation of the exhibited preference for unequal recombination in a small region, for example the 3.0-kb hotspot in SoS is only ~5% in size of the total PLCR-B, other role-playing factors seem likely. A specific chromatin structure was hypothesized to be such a determining factor (17,18). A susceptible conformation would possibly have increased accessibility for the double-strand break and repair machinery and could consequently predispose for a hotspot location (17).

The human chromatin is organized in around 60,000 loop domains which are periodically attached at their base to a supporting skeleton, the so called nuclear scaffold or matrix (20). This compartmentalization of the genome has an important regulatory function in gene expression, DNA replication and recombination (20,36). Since S/MARs are essentially recombinogenic unpairing regions, a strong correlation has been found between two basically

different algorithms, the S/MAR-wiz and WebSIDD, with the latter detecting stress-induced destabilized unwound DNA (37). In general, results of *in silico* analysis should be considered carefully. However, a good correlation for the used programs was already confirmed with the results of *in vitro* experiments (38).

As many as 74% (23/31) of the breakpoints in the *M*ixed *L*ineage *L*eukemia (*MLL*) gene in *de novo* leukemia patients were mapped to a breakpoint cluster region located between S/MARs (20). Furthermore, a clustering of breakpoints of t(4;11) translocations in the human *MLL* and *A*LL-1 *F*used chromosome *4* (*AF4*) genes was also found to be located outside high-affinity S/MARs, but with flanking S/MARs in the vicinity (39). The hotspots for SoS, NF1, SMS/dup(17)(p11.2p11.2), and uncommon deletions in SMS as investigated with these programs in this study, showed a similar pattern of stabilized DNA duplex regions, located between destabilized regions with a coinciding higher S/MAR probability. In contrast, the four non-hotspot-related breakpoints were found in or at the border of highly destabilized DNA with an increased S/MAR potential. The patterns of the hotspot for deletions and reciprocal duplications in HNPP/CMT1A were not in complete correlation. The hotspot still seemed to be located within a stabilized DNA helix, but also a S/MAR potential of ~0.10 was found. As the S/MAR potential in the distal LCR was close to zero (data not the shown), the meaning of such a slightly increased potential remains to be investigated. The differences in DNA destabilization profiles and in frequency of occurrence between breakpoints located in and outside the SoS hotspot seem to support the idea that the center for recombination is located in stabilized DNA regions and that regions with strand separation potential (i.e. S/MARs) are likely to function as mediators (20). However, it should be noted that the previous data is based upon somatic events in leukemia patients with translocations between different chromosomes (20,39). Currently only a limited number of genomic disorders could be used for analysis. Therefore, future identification and analysis of other breakpoint clusters and non-hotspot-related breakpoints mediated by NAHR, will possibly determine whether this analysis could be used in combination with other hotspot characteristics to predict possible recombination hotspot locations within LCRs.

In conclusion, the first identification of four non-hotspot-related breakpoints in SoS in comparison with the SoS and other recombination hotspots indicates that DNA duplex stabilization and specific chromatin organisation might play a role in predisposition for recombination hotspot locations of genomic disorders.

Acknowledgements

We kindly express our gratitude to the patients, their parents and the referring physicians for their cooperation. Furthermore we thank Ms. Tamae Hanai and Ms. Yasuko Noguchi for their excellent technical assistance. This study is supported by the Japan Science and Technology Agency (CREST) and the International Consortium for Medical Care of Hibakusha and Radiation Life Science, The 21st Century Center of Excellence (COE).

References

1. Cole TRP, Hughes HE. Sotos syndrome: a study of the diagnostic criteria and natural history. J Med Genet 1994; 31: 20-32
2. Kurotaki N, Imaizumi K, Harada N, Masuno M, Kondoh T, Nagai T, et al. Haploinsufficiency of *NSD1* causes Sotos syndrome. Nat Genet 2002; 30: 365-366
3. Kurotaki N, Harada N, Shimokawa O, Miyake N, Kawame H, Uetake K, et al. Fifty microdeletions among 112 cases of Sotos syndrome: low copy repeats possibly mediate the common deletion. Hum Mutat 2003; 22: 378-387
4. Douglas J, Hanks S, Temple IK, Davies S, Murray A, Upadhyaya M, et al. *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 2003; 72: 132-143
5. Kamimura J, Endo Y, Kurotaki N, Kinoshita A, Miyake N, Shimokawa O, et al. Identification of eight novel NSD1 mutations in Sotos syndrome. J Med Genet 2003; 40: e126
6. Nagai T, Matsumoto N, Kurotaki N, Harada N, Niikawa N, Ogata T, et al. Sotos syndrome and haploinsufficiency of *NSD1*: clinical features of intragenic mutations and submicroscopic deletions. J Med Genet 2003; 40: 285-289
7. Rio M, Clech L, Amiel J, Faivre L, Lyonnet S, Le Merrer M, et al. Spectrum of *NSD1* mutations in Sotos and Weaver syndromes. J Med Genet 2003; 40: 436-440
8. Turkmen S, Gillessen-Kaesbach G, Meinecke P, Albrecht B, Neumann LM, Hesse V, et al. Mutations in NSD1 are responsible for Sotos syndrome, but are not a frequent finding in other overgrowth phenotypes. Eur J Hum Genet 2003; 11: 858-865
9. de Boer L, Kant SG, Karperien M, van Beers L, Tjon J, Vink GR, et al. Genotype-phenotype correlation in patients suspected of having sotos syndrome. Horm Res 2004; 62: 197-207
10. Visser R, Matsumoto N. Genetics of Sotos syndrome. Curr Opin Pediatr 2003; 15: 598-606
11. Visser R, Shimokawa O, Harada N, Kinoshita A, Ohta T, Niikawa N, et al. Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion. Am J Hum Genet 2005; 76: 52-67 [Epub 2004 November 16]
12. Kurotaki N, Stankiewicz P, Wakui K, Niikawa N, Lupski JR. Sotos syndrome common deletion is mediated by directly oriented subunits within inverted Sos-REP low-copy repeats. Hum Mol Genet 2005; 14: 535-542 [published Online First: 7 January 2005]
13. Reiter LT, Murakami T, Koeuth T, Pentao L, Muzny DM, Gibbs RA, et al. A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element. Nat Genet 1996; 12: 288-297
14. Lopez-Correa C, Dorschner M, Brems H, Lazaro C, Clementi M, Upadhyaya M, et al. Recombination hotspot in NF1 microdeletion patients. Hum Mol Genet 2001; 10: 1387-1392

15. Bi W, Park SS, Shaw CJ, Withers MA, Patel PI, Lupski JR. Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2. Am J Hum Genet 2003; 73: 1302-1315
16. Shaw CJ, Lupski JR. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet 2004; 13 Suppl 1: R57-64
17. Lupski JR. Hotspots of homologous recombination in the human genome: not all homologous sequences are equal. Genome Biol 2004; 5: 242
18. Bacolla A, Jaworski A, Larson JE, Jakupciak JP, Chuzhanova N, Abeysinghe SS, et al. Breakpoints of gross deletions coincide with non-B DNA conformations. Proc Natl Acad Sci U S A 2004; 101: 14162-14167
19. Albrecht P, Bode J, Buiting K, Prashanth AK, Lohmann DR. Recurrent deletion of a region containing exon 24 of the RB1 gene caused by non-homologous recombination between a LINE-1HS and MER21B element. J Med Genet 2004; 41: e122
20. Bode J, Benham C, Ernst E, Knopp A, Marschalek R, Strick R, et al. Fatal connections: when DNA ends meet on the nuclear matrix. J Cell Biochem Suppl 2000; Suppl 35: 3-22
21. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000; 132: 365-386
22. Estivill X, Cheung J, Pujana MA, Nakabayashi K, Scherer SW, Tsui LC. Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome. Hum Mol Genet 2002; 11: 1987-1995
23. Reiter LT, Hastings PJ, Nelis E, De Jonghe P, Van Broeckhoven C, Lupski JR. Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. Am J Hum Genet 1998; 62: 1023-1033
24. Lopes J, Ravise N, Vandenberghe A, Palau F, Ionasescu V, Mayer M, et al. Fine mapping of de novo CMT1A and HNPP rearrangements within CMT1A-REPs evidences two distinct sex-dependent mechanisms and candidate sequences involved in recombination. Hum Mol Genet 1998; 7: 141-148
25. Lopes J, Tardieu S, Silander K, Blair I, Vandenberghe A, Palau F, et al. Homologous DNA exchanges in humans can be explained by the yeast double-strand break repair model: a study of 17p11.2 rearrangements associated with CMT1A and HNPP. Hum Mol Genet 1999; 8: 2285-2292
26. Shaw CJ, Withers MA, Lupski JR. Uncommon deletions of the Smith-Magenis syndrome region can be recurrent when alternate low-copy repeats act as homologous recombination substrates. Am J Hum Genet 2004; 75: 75-81
27. Bi C, Benham CJ. WebSIDD: server for predicting stress-induced duplex destabilized (SIDD) sites in superhelical DNA. Bioinformatics 2004; 20: 1477-1479
28. Singh GB, Kramer JA, Krawetz SA. Mathematical model to predict regions of chromatin attachment to the nuclear matrix. Nucleic Acids Res 1997; 25: 1419-1425

29. Miyake N, Kurotaki N, Sugawara H, Shimokawa O, Harada N, Kondoh T, et al. Preferential paternal origin of microdeletions caused by prezygotic chromosome or chromatid rearrangements in Sotos syndrome. Am J Hum Genet 2003; 72: 1331-1337
30. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, et al. Human l1 retrotransposition is associated with genetic instability in vivo. Cell 2002; 110: 327-338
31. Kazazian HH, Jr., Goodier JL. LINE drive: retrotransposition and genome instability. Cell 2002; 110: 277-280
32. Deininger PL, Batzer MA. Alu repeats and human disease. Mol Genet Metab 1999; 67: 183-193
33. Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 1998; 14: 417-422
34. Waldman AS, Liskay RM. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol Cell Biol 1988; 8: 5350-5357
35. Inoue K, Lupski JR. Molecular mechanisms for genomic disorders. Annu Rev Genomics Hum Genet 2002; 3: 199-242
36. Bode J, Goetze S, Heng H, Krawetz SA, Benham C. From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. Chromosome Res 2003; 11: 435-445
37. Benham C, Kohwi-Shigematsu T, Bode J. Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions. J Mol Biol 1997; 274: 181-196
38. Goetze S, Gluch A, Benham C, Bode J. Computational and in vitro analysis of destabilized DNA regions in the interferon gene cluster: potential of predicting functional gene domains. Biochemistry 2003; 42: 154-166
39. Hensel JP, Gillert E, Fey GH, Marschalek R. Breakpoints of t(4;11) translocations in the human MLL and AF4 genes in ALL patients are preferentially clustered outside of high-affinity matrix attachment regions. J Cell Biochem 2001; 82: 299-309
40. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. Trends Genet 2002; 18: 74-82