

# Making of fusion genes in cancer: An in-silico study of mechanism of chromosomal translocations

Kiran Lalwani <sup>1</sup>, Shivani Sheth <sup>1</sup>, Inayatullah Sheikh <sup>2</sup>, Afzal Ansari <sup>2</sup>, Fulesh Kunwar <sup>1</sup>, Jayshankar Das <sup>2</sup>,  
Sonal R Bakshi <sup>Corresp. 1</sup>

<sup>1</sup> Institute of science, Nirma University, Ahmedabad, Gujarat, India

<sup>2</sup> Department of Science and Technology, Gujarat Biotechnology Research Centre, Government of Gujarat, Gandhinagar, Gujarat, India

Corresponding Author: Sonal R Bakshi

Email address: sonal.bakshi@nirmauni.ac.in

Chromosomal translocations involve exchange of genetic material between non-homologous chromosomes leading to the formation of a fusion gene with altered function. The clinical consequences of non-random and recurrent chromosomal translocations have been so well understood in carcinogenesis that they serve as diagnostic and prognostic markers and also help in therapy decisions, mainly in leukemia and lymphoma. However, the molecular mechanisms underlying these recurrent genetic exchanges are yet to be understood. Various approaches employed include the extent of the vicinity of the partner chromosomes in the nucleus, DNA sequences at the breakpoints, etc. The present study addresses the stability of DNA sequences at the breakpoint regions using in-silico approach in terms of physicochemical properties such as; AT%, flexibility, melting temperature, enthalpy, entropy, stacking energy and free energy. Changes in these properties may lead to instability of DNA which could affect gene expression in particular and genome organization in general. Our study indicates that the fusion sequences are comparatively more unstable and hence, more prone to breakage. Current study along with others could lead to developing a model for predicting breakage prone genomic regions using this novel in-silico approach.

1 **Making of Fusion Genes in Cancer: An in silico study of mechanism of chromosomal**  
2 **translocations**

3

4

5 Kiran Lalwani<sup>1</sup> ; Shivani Sheth<sup>1</sup> ; Inayatullah Shaikh<sup>2</sup> ; Afsal Ansari<sup>2</sup> ;

6 Fulesh Patel<sup>1</sup> ; Dr. Jayshankar Das<sup>2</sup> ; Dr. Sonal R Bakshi<sup>1</sup>

7

8

9 <sup>1</sup>Institute of Science, Nirma University, Ahmedabad, Gujarat, India

10 <sup>2</sup>Department of Science and Technology, Gujarat Biotechnology Research Center, Government  
11 of Gujarat, Gandhinagar, Gujarat, India

12

13

14 Corresponding Author: Dr. Sonal R. Bakshi,

15 Assistant Professor, Institute of Science, Nirma University,

16 Sarkhej- Gandhinagar Highway,

17 Ahmedabad- 382481, Gujarat, India,

18 Phone- 079-30642755, Mobile- 9426417807, Fax- +9102717 24191

19 Email- sonal.bakshi@nirmauni.ac.in

20

21

22

**23 Abstract**

24

25 Chromosomal translocations involve exchange of genetic material between non- homologous  
26 chromosomes leading to the formation of a fusion gene with altered function. The clinical  
27 consequences of non-random and recurrent chromosomal translocations have been so well  
28 understood in carcinogenesis that they serve as diagnostic and prognostic markers and also help  
29 in therapy decisions, mainly in leukemia and lymphoma. However, the molecular mechanisms  
30 underlying these recurrent genetic exchanges are yet to be understood. Various approaches  
31 employed include the extent of the vicinity of the partner chromosomes in the nucleus, DNA  
32 sequences at the breakpoints, etc. The present study addresses the stability of DNA sequences at  
33 the breakpoint regions using in-silico approach in terms of physicochemical properties such as;  
34 AT%, flexibility, melting temperature, enthalpy, entropy, stacking energy and free energy.  
35 Changes in these properties may lead to instability of DNA which could affect gene expression  
36 in particular and genome organization in general. Our study indicates that the fusion sequences  
37 are comparatively more unstable and hence, more prone to breakage. Current study along with  
38 others could lead to developing a model for predicting breakage prone genomic regions using  
39 this novel in-silico approach.

40

41

42 Keywords: Chromosomal translocations, carcinogenesis, instability of DNA, in-silico approach,  
43 physicochemical properties, fusion sequences, gene expression, prone to breakage

44

45

**46 Introduction**

47

48 Chromosomal translocations involve partial exchange of genetic material among non-  
49 homologous chromosomes<sup>1</sup>. The resulting chimeric fusion transcripts are reported to code for  
50 proteins that impart selective growth advantage thus triggering carcinogenesis<sup>2</sup>. The fusion  
51 transcripts and proteins are also characteristic markers of diagnosis and therapy<sup>3</sup>. Various  
52 leukemia and lymphoma subtypes are based on the type of chromosome translocation as per the  
53 WHO-2008 classification which is more clinically relevant as compared to the FAB  
54 classification. Some of the examples of significant oncogenic chromosomal fusions are t(9;22),  
55 t(15;17), t(8;21) and t(16;16). These translocations result in the formation of unique chimeric  
56 proteins which activate a number of signal pathways leading to malignant transformation by  
57 interfering with the basic cellular process, such as control of cell proliferation and differentiation,  
58 adhesion and cell survival<sup>4, 18</sup>.

59

60

61 The fusion t(9;22) occurs between coding sequence of BCR gene on chromosome 22 with that of  
62 ABL1 gene on chromosome 9, hence result in derivative chromosome viz. Philadelphia

63 chromosome. The BCR-ABL fusion protein is a well-defined marker in Chronic Myeloid  
64 Leukemia<sup>5</sup>.

65  
66

67 The molecular analysis of Acute Promyelocytic Leukemia (APL) breakpoint has shown the  
68 involvement of RARA gene on chromosome 17 with PML gene on chromosome 15. Two  
69 reciprocal fusion products are formed, namely, PML/RARA and RARA/PML. Out of these,  
70 PML/RARA is implicated in leukemogenesis and may interfere with the normal functions of  
71 PML and/or RARA. APL is FAB M3 subtype of AML<sup>6,7</sup>.

72  
73

74 The t(8;21) translocation is one of the most frequent chromosome abnormalities in AML and is  
75 morphologically associated with the FAB-M2 subtype of AML<sup>5,7</sup>. It results in an in-frame fusion  
76 of two genes, AML1 and ETO. Due to translocation, the AML1 gene (also called RUNX1) on  
77 chromosome 21 fuses with the ETO gene (also referred to as the RUNX1T1) on chromosome 8.  
78 The t(16;16) is a rare chromosomal abnormality in AML, also classified as AML- M4Eo  
79 according to FAB classification<sup>9</sup>. It involves the fusion of the CBFβ (Core-Binding Factor Beta  
80 Subunit) gene at 16q22 with the smooth muscle myosin heavy chain (MYH11) at 16p13. This  
81 results in a generation of the CBFβ-MYH11 fusion protein.

82  
83

84 The occurrence of recurrent translocations can be hypothetically explained in various ways. We  
85 have earlier reported the study of physical proximity of chromosomes undergoing  
86 translocation<sup>10</sup>. In the present study, we address the properties of the breakpoint region of the  
87 chromosomes taking part in translocation as a possible factor playing a role in recurrent breakage  
88 and reunion of two chromosomal sites. The chromosomal sites involved in translocation are  
89 likely to play a role in translocation owing to their physicochemical properties such as AT%,  
90 flexibility, melting temperature, enthalpy, entropy, stacking energy and free energy. This was  
91 assessed along with a reference sequence that is flanking the breakpoint region. The information  
92 regarding physicochemical properties of gene sequences taking part in translocation could  
93 provide useful insights on the precise investigations of the mechanism of leukemogenesis.

94  
95

96 The molecular and stochastic mechanism underlying the chromosomal translocations is still  
97 speculative and requires further studies involving varied chromosomal regions. The  
98 chromosomal translocations in leukemia are ideal candidates for such studies owing to the non-  
99 random and well-characterized breakpoints of the gene pairs that provide precise sequences for  
100 investigation. The availability of various computational approaches has generated a number of  
101 in-silico studies<sup>11</sup>. In one such study, a unique computational multi-parametric bio-informatics  
102 methodology was employed for analysis of genomic sequences of breakpoint coordinates, which

103 provided some perspective of molecular mechanisms of DNA strand break during translocation  
104 leading to epithelial carcinogenesis. This study led to the conceptualization of an algorithmic  
105 program which may predict possible breakpoint in any given sequence of human genome<sup>12</sup>.  
106 Appertaining to this prior study, we attempt to analyze the sequence properties of breakpoint  
107 regions of chromosome pairs taking part in translocation in selected four cases of leukemia  
108 subtypes that are known to be recurrent and also well characterized in terms of breakpoints.  
109 Three of the translocation pairs are non-homologous chromosomes (9/22, 8/21, and 15/17);  
110 whereas one translocation involves two different regions on the same chromosome on two  
111 different arms (16p/16q). Unlike the first three cases, the fourth case is unique as in this case the  
112 physical proximity of fusion partners is ascertained being on the same chromosome.

113

114

115 In the present study, we carried out in-silico analysis of DNA sequences present at the  
116 chromosome translocation breakpoints.

117

118

## 119 **Materials and methods**

120

### 121 *Data mining*

122

123 The in-silico analysis involved the use of various computational tools so as to obtain the fusion  
124 gene sequences and to acquire the physicochemical properties of the sequences flanking them  
125 (Figure 1).

126

127

128 Retrieval of the fusion sequences of BCR-ABL, PML-RARA, AML-ETO, and CBFβ- MYH11,  
129 which are involved in the formation of fusion products, was done using TICdb  
130 ([www.unav.es/genetica/TICdb](http://www.unav.es/genetica/TICdb)). The retrieved fusion sequences were entered in BLAT (BLAST-  
131 like Alignment Tool) query box (<http://genome.ucsc.edu/>). From the BLAT search results, the  
132 ones showing maximum identity with reference sequence were selected. Maximum identity is  
133 the extent of similarity that the fusion sequence shows with the reference sequence of the human  
134 genome considered in the BLAT tool.

135

136

137 The UCSC genome browser was further used to retrieve 500 base pairs upstream and  
138 downstream (a 1000 bp sequence) of the breakpoints of the fusion gene product for both the  
139 partner chromosomes. Similarly, 50 base pairs upstream and downstream (a 100 bp sequence) of  
140 each gene were retrieved. In order to compare the sequences involved in the formation of fusion  
141 genes, exonic sequences of 1000 as well as 100 bp length were similarly retrieved for BCR,  
142 ABL, PML, RARA, AML1, ETO, MYH11, and CBFβ these were taken as 'control'.

143

144

145 *Physicochemical properties*

146

147 The physicochemical properties of breakpoint region are assumed to be responsible for the easy  
148 breakage and reunion with the non-homologous partner sequence; hence we hypothesized that  
149 the genomic regions neighboring the breakpoint of fusion sequences would be different in terms  
150 of physicochemical properties and hence selected 50 bps downstream and upstream of  
151 breakpoints of fusion sequence as internal controls. Similarly, in order to increase the sample  
152 size, a similar exercise was also done using long stretches of genomic sequence i.e. 500 bp  
153 upstream and downstream of breakpoints of fusion sequence as internal controls.

154

155

156 For the evaluation of the physicochemical properties of the fusion sequences obtained for all  
157 eight genes, including their exonic sequences, DiProGB (<http://diprogb.flileibniz.de/>) was used.  
158 The properties selected for evaluation were AT%, flexibility, melting temperature, enthalpy,  
159 entropy, free energy and stacking energy.

160

161

## 162 **Results and Discussion**

163

164 The total number of fusion sequences obtained for all four translocation pairs varied. For BCR-  
165 ABL fusion sequences obtained was 16, out of which those fusion sequences showing maximum  
166 identity were 12. While the total number of sequences obtained for PML-RARA was 26, out of  
167 which those with maximum identity were 18. For AML1-ETO 17 out of 37 sequences showed  
168 maximum sequence identity. And for CFBF-MYH11 all the 7 sequences showed sequence  
169 maximum identity.

170

171

172 Further comparison of values of the different physicochemical properties of the sequences  
173 involved in translocation and the controls was carried out as shown in table 1 and 2. The values  
174 shown are an average of all the 12, 18, 17 and 7 fusion sequences for the eight genes as well as  
175 their controls. Our data as shown in the tables and supplementary material suggest that the  
176 breakpoints of fusion sequence and controls had different AT% and related parameters as  
177 compared when compared with the two different base-pair lengthed internal controls. The extent  
178 of the difference in AT % and hence the related parameters was comparable for both, the 100 and  
179 1000 bps internal controls.

180

181

182 Several physicochemical parameters such as melting temperature, stacking energy, free energy

183 etc. are measured as these are known to stabilize the DNA<sup>13</sup>. An evaluation of these factors can  
184 provide an explanation of various properties and functional interactions of DNA. These diverse  
185 factors that stabilize DNA act in concert to protect it from random breakage. If these factors  
186 remain consistent, they could prevent the genome from damage, but it is known that DNA is a  
187 dynamic entity and the confounding factors do show variation leading to DNA instability. The  
188 instability of DNA could affect gene expression in particular and genome organization in  
189 general<sup>13</sup>. Evaluation of the comprehensive data regarding physicochemical features of the fusion  
190 gene partners, when compared to the normal gene exons, showed variations in all the parameters  
191 (table 3).

192

193

194 Our results demonstrate that AT% is in direct proportion to the flexibility. The flexibility of  
195 DNA is important for its biological functions such as gene expression, gene regulation etc. It is  
196 influenced by several factors like sequence arrangement, temperature etc. It is well known that  
197 DNA flexibility depends on its sequence, which directly affects its stability<sup>14, 15</sup>. Numerous  
198 previous studies have shown that different sequence arrangements can greatly influence the  
199 stability of DNA. The AT-rich sequences are more flexible than the GC rich sequences. This  
200 increase in the AT%, as well as the flexibility of the sequences involved in the formation of  
201 fusion products in comparison to the controls, could indicate an increase in instability.

202

203

204 Melting temperature indicates the temperature at which half of the double-stranded DNA gets  
205 converted to single-stranded DNA. Since A and T are linked with each other with double  
206 hydrogen bonds, the DNA having higher AT content will have a relatively lower melting  
207 temperature. Also, the temperature can directly determine the state of the structure of DNA and  
208 thus greatly influence its flexibility<sup>14, 15</sup>. Our study correlates this decrease in the melting  
209 temperatures with an increase in flexibility and AT content for the fusion sequences.

210

211

212 As the hydrogen bonding between A and T would not alone lead to DNA instability, the stacking  
213 energy of the DNA could be taken into consideration. The stacking energy of the AT bases is  
214 generally weaker than the GC base pairs. Due to this, the AT-rich regions have a lower melting  
215 temperature as compared to the GC rich regions. The stacking energy hence can determine the  
216 DNA stability and its disruption can contribute to decrease in melting temperature and also  
217 change the thermal energy of the molecule. In our study, there was a significant decline in the  
218 stacking energy of the fusion sequences when compared to the stacking energy of the control<sup>16</sup>.

219

220

221 Thermodynamics studies suggest that processes like the formation of double helix will happen  
222 spontaneously only if they result in a net increase in entropy<sup>17</sup>. There are reports which indicate

223 that more negative the free energy is, more is the stability of the DNA. On the basis of the  
224 relationship between standard free energy change ( $\Delta G$ ), enthalpy ( $\Delta H^\circ$ ) and entropy ( $\Delta S^\circ$ ), given  
225 by the equation,  $\Delta G = \Delta H^\circ - T\Delta S^\circ$ , we assume that a net decrease in free energy may result  
226 because of a decrease in enthalpy and entropy individually. Our results support this assumption;  
227 however additional assessment on the thermodynamic properties of DNA in relation to its  
228 structure needs to be carried out.

229

230

## 231 **Conclusion**

232

233 The mechanism underlying chromosomal translocations remains under speculation and requires  
234 advanced investigations. Based on our in-silico studies, the physicochemical parameters that are  
235 known to affect DNA stability viz., AT%, flexibility, melting temperature, stacking energy,  
236 enthalpy, entropy and free energy used for comparison of the sequences involved in formation of  
237 fusion products with the controls manifested instability of the sequences flanking the breakpoint  
238 regions. The overall decrease in AT% and flexibility and a gross increase in remaining properties  
239 indicated that these sequences were found to be more prone to breakage. Extensive work carried  
240 out in the direction of this novel approach could lead to the development of an in-silico  
241 predictive model that aids in finding a particular genomic area or region that is prone to frequent  
242 breakage.

243

244

## 245 **Study Highlights**

246

247 What is the current knowledge?

- 248 • Recurrent chromosome translocations leading to fusion genes are known to be pathogenic  
249 in cancer
- 250 • The mechanism underlying non-random involvement of breakpoints in gene pairs is not  
251 well understood despite various approaches reported in terms of the vicinity, DNA  
252 packaging, etc.

253

254 What is new here?

- 255 • The physicochemical properties of breakpoints in gene pairs participating in fusion are  
256 studied for instability favoring rearrangement
- 257 • The in-silico approach to study this mechanism in case of hematological malignancies  
258 has not been reported earlier

259

260

## 261 **References**

262



- 263 1. Nickoloff, Jac A., et al. "Mechanisms of Leukemia Translocations." *Current Opinion in*  
264 *Hematology*, vol. 15, no. 4, July 2008, pp. 338–45. PubMed Central,  
265 doi:10.1097/MOH.0b013e328302f711
- 266 2. Zheng, Jie. "Oncogenic Chromosomal Translocations and Human Cancer (Review)."  
267 *Oncology Reports*, vol. 30, no. 5, Nov. 2013, pp. 2011–19.
- 268 3. Nambiar, Mridula, et al. "Chromosomal Translocations in Cancer." *Biochimica Et*  
269 *Biophysica Acta*, vol. 1786, no. 2, Dec. 2008, pp. 139–52. PubMed,  
270 doi:10.1016/j.bbcan.2008.07.005.
- 271 4. Salesse, Stephanie, and Catherine M. Verfaillie. "BCR/ABL: From Molecular  
272 Mechanisms of Leukemia Induction to Treatment of Chronic Myelogenous Leukemia."  
273 *Oncogene*, vol. 21, no. 56, Dec. 2002, pp. 8547–59. PubMed,  
274 doi:10.1038/sj.onc.1206082.
- 275 5. O'Connor Clare. "Human chromosome translocations and cancer" *Nature Education* 2008  
276 1(1):56
- 277 6. Tallman MS, Altman JK, Curative strategies in acute promyelocytic leukemia,  
278 *Hematology Am Soc Hematol Educ Program*. 2008:391-9
- 279 7. C C Coombs, M Tavakkoli and M S Tallman, Acute promyelocytic leukemia: where did  
280 we start, where are we now, and the future, *Blood Cancer Journal* (2015), 17 April 2015
- 281 8. Reikvam, Håkon, et al. "Acute Myeloid Leukemia with the t(8;21) Translocation:  
282 Clinical Consequences and Biological Implications." *BioMed Research International*,  
283 2011, doi:10.1155/2011/104631.
- 284 9. Adriaansen, H., te Boekhorst, P., Hagemeijer, A., van der Schoot, C., Delwel, H., & van  
285 Dongen, J. (1993). Acute myeloid leukemia M4 with bone marrow eosinophilia (M4Eo)  
286 and inv(16)(p13q22) exhibits a specific immunophenotype with CD2  
287 expression. *Blood*, 81(11),3043-3051
- 288 10. Fulesh Kunwar, Kiran Tandel and Sonal R. Bakshi, Recurrent chromosomal  
289 translocations: Is proximity a rule? *Indian Journal for Experimental Biology (IJEB)*,  
290 2017, Vol. 55 (1) 15-20
- 291 11. Murray, David, et al. "In-silico Gene Expression Analysis – an Overview." *Molecular*  
292 *Cancer*, vol. 6, Aug. 2007, p. 50. PubMed Central, doi:10.1186/1476-4598-6-50.
- 293 12. Daga, Aditi, et al. "Characterization of Chromosomal Translocation Breakpoint  
294 Sequences in Solid Tumours: 'An In-silico Analysis.'" *The Open Medical Informatics*

295 Journal, vol. 9, Apr. 2015, pp. 1–8. PubMed Central, doi:  
296 10.2174/1874431101509010001.

297 13. Khandelwal, Garima, and Jayaram Bhyravabhotla. “A Phenomenological Model for  
298 Predicting Melting Temperatures of DNA Sequences.” PLoS ONE, vol. 5, no. 8, Aug.  
299 2010. PubMed Central, doi:10.1371/journal.pone.0012433.

300 14. Puri, Dinesh. Textbook of Medical Biochemistry. Elsevier Health Sciences, 2014.

301 15. Bao, Lei, et al. “Flexibility of Nucleic Acids: From DNA to RNA.” ArXiv:1509.06450  
302 [Cond-Mat, Physics:Physics], Sept. 2015. arXiv.org, <http://arxiv.org/abs/1509.06450>

303 16. Laurence A. Moran, DNA Denaturation and Renaturation and the Role of Hydrogen  
304 Bonds and Stacking Interactions, Sandwalk, Saturday, December 29, 2007

305 17. Petruska, J., and M. F. Goodman. “Enthalpy-Entropy Compensation in DNA Melting  
306 Thermodynamics.” The Journal of Biological Chemistry, vol. 270, no. 2, Jan. 1995, pp.  
307 746–50.

308 18. Frazer, Rachel, et al. “Chronic Myeloid Leukaemia in The 21st Century.” The Ulster  
309 Medical Journal, vol. 76, no. 1, Jan. 2007, pp. 8–17.

310

311

312

313

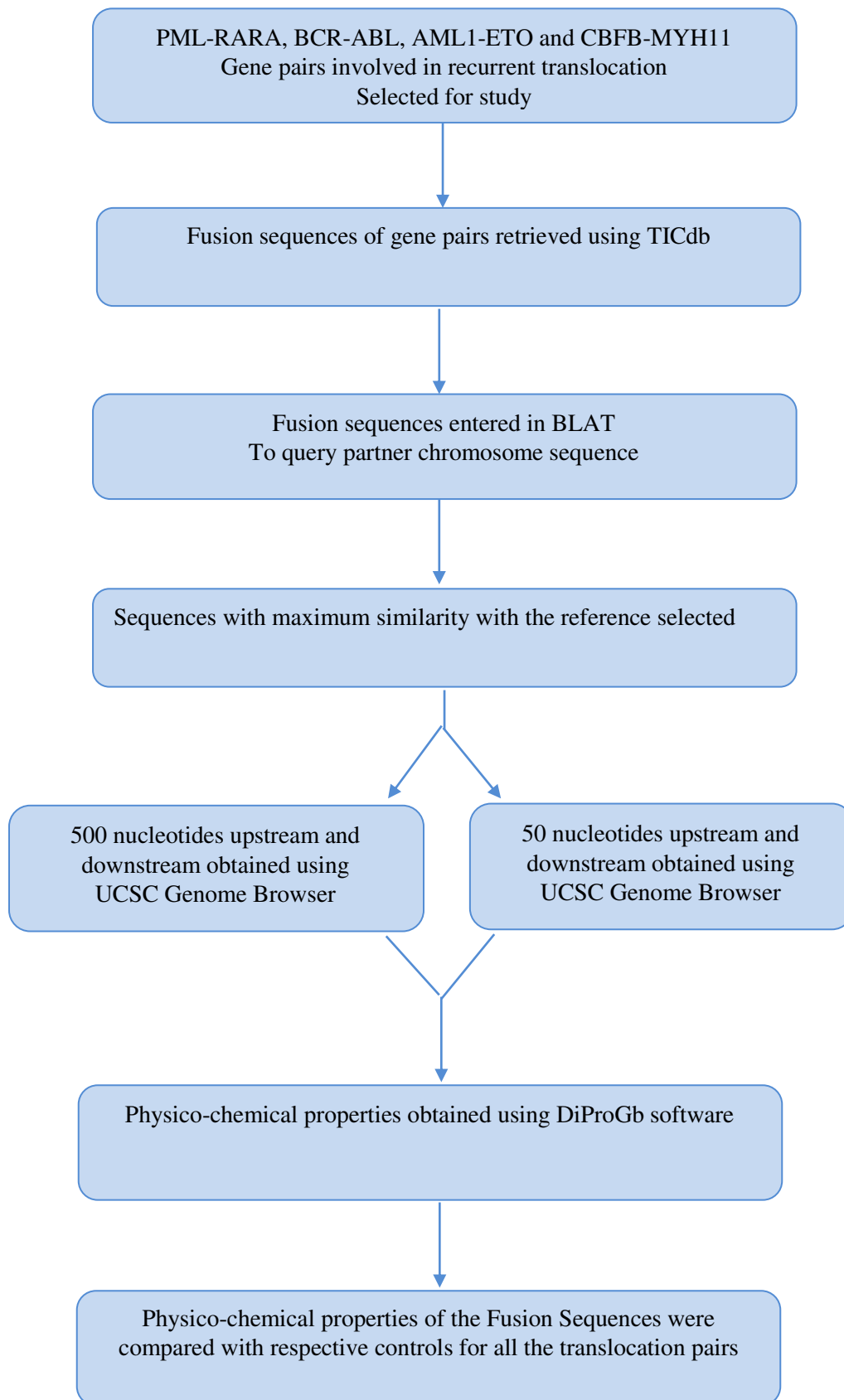
314

315

316

**Figure 1** (on next page)

Flowchart depicting data retrieval of 500 and 50 base pairs upstream and downstream to the breakpoint flanking region of fusion gene pairs



**Figure 1:** Flowchart depicting data retrieval of 500 and 50 base pairs upstream and downstream to the breakpoint flanking region of fusion gene pairs

**Table 1** (on next page)

Analysis of physico-chemical parameters/AT% within 1000 bps around breakpoint for Fusion Sequence (FS) and Control (C)

Average of all physico-chemical parameters/AT% within 1000 bps around breakpoint for Fusion Sequence (FS) and Control (C) for BCR, ABL, PML, RARA, AML1, ETO, CBFB, and MYH11

**Table 1:** Analysis of physico-chemical parameters/AT% within 1000 bps around breakpoint for Fusion Sequence (FS) and Control (C)

Gene	Sequence	AT (%)	Flexibility (kJ/mol A2)	Melting Temperature (°C)	Enthalpy (kcal/mol)	Entropy (cal/mol/K)	Stacking Energy (kcal/mol)	Free Energy (kcal/mol)
ABL	1000 FS	58.525	8.47	69.826	-8.318	-21.936	-7.567	-1.096
	1000 C	50.15	7.937	73.121	-8.558	-22.427	-7.947	-1.174
BCR	1000 FS	44.375	7.523	75.125	-8.7192	-22.779	-8.162	-1.223
	1000 C	44.2	7.47	75.451	-8.726	-22.773	-8.197	-1.231
PML	1000 FS	46.922	7.623	74.115	-8.593	-22.472	-8.05	-1.201
	1000 C	44.442	7.442	75.336	-8.687	-22.658	-8.195	-1.23
RARA	1000 FS	41.5	7.482	75.854	-8.808	-22.98	-8.166	-1.248
	1000 C	40.5142	7.265	76.785	-8.8485	-23.04	-8.32	-1.266
AML1	1000 FS	64.952	8.851	66.781	-8.06	-21.371	-7.25	-1.024
	1000 C	60.95	8.548	68.883	-8.211	-21.7	-7.5	-1.071
ETO	1000 FS	55.08	8.142	70.9	-8.358	-21.9	-7.707	-1.123
	1000 C	55.85	8.251	70.51	-8.353	-22.004	-7.66	-1.11
CBFB	1000 FS	64.95	8.85	66.781	-8.058	-21.371	-7.249	-1.024
	1000 C	60.96	8.548	68.88	-8.211	-21.7	-7.493	-1.071
MYH11	1000 FS	55.082	8.142	70.89	-8.3584	-21.99	-7.707	-1.1236
	1000 C	55.85	8.251	70.51	-8.353	-22.004	-7.657	-1.113



**Table 2** (on next page)

Analysis of physico-chemical parameters/AT% within 100 bps around breakpoint for Fusion Sequence (FS) and Control (C)

Average of all physico-chemical parameters/AT% within 100 bps around breakpoint for Fusion Sequence (FS) and Control (C) for BCR, ABL, PML, RARA, AML1, ETO, CFBF and MYH11



**Table 2:** Analysis of physico-chemical parameters/AT% within 100 bps around breakpoint for Fusion Sequence (FS) and Control (C)

Gene	Sequence	AT( %)	Flexibility (kJ/molA <sup>2</sup> )	Melting Temperature (°C)	Enthalpy (kcal/mol)	Entropy (cal/mol/K)	Stacking Energy (kcal/mol)	Free Energy (kcal/mol)
ABL	100 FS	53.34	7.646	75.407	-8.768	-22.908	-8.202	-1.227
	100 C	44.1	7.397	75.582	-8.73	-22.78	-8.232	-1.234
BCR	100 FS	44.917	8.491	70.812	-8.378	-22.034	-7.676	-1.118
	100 C	43.2	7.57	75.63	-8.67	-22.6	-8.302	-1.227
PML	100 FS	45.56	7.478	75.155	-8.758	-22.88	-8.115	-1.229
	100 C	38.85	7.097	78.3	-8.97	-23.307	-8.56	-1.296
RARA	100 FS	43.78	7.682	74.912	-8.686	-22.683	-8.101	-1.225
	100 C	42.86	7.406	76.35	-8.79	-22.9	-8.316	-1.254
AML1	100 FS	66.94	8.983	65.833	-7.991	-21.234	-7.141	-1.003
	100 C	47.89	7.77	74.216	-8.632	-22.58	-8.091	-1.19
ETO	100 FS	55.059	8.0922	71.092	-8.348	-21.96	-7.76	-1.127
	100 C	43.25	7.28	76.28	-8.8	-22.937	-8.35	-1.246
CBFB	100 FS	52.857	8.207	72.095	-8.452	-22.205	-7.87	-1.146
	100 C	44.67	7.866	76.1	-8.81	-22.97	-8.362	-1.237
MYH11	100 FS	42.71	7.824	75.8	-8.771	-22.844	-8.247	-1.24
	100 C	37	7.38	77.29	-8.856	-22.97	-8.421	-1.269



**Table 3** (on next page)

Overall comparison between Fusion sequences and Control

The table shows a general comparison of all the physico-chemical properties of all four fusion pairs with their respective exons for both 100bps as well as for 1000 bps. There is an overall increase in flexibility and AT content in the fusion sequences while other parameters like Stacking energy, free energy, entropy, enthalpy and melting temperatures show an overall decrease when compared to the control sequences.

Parameters	1000 bps		100 bps	
	Fusion sequence	Control	Fusion sequence	Control
AT (%)	↑	↓	↑	↓
Flexibility (kJ/molA <sup>2</sup> )	↑	↓	↑	↓
Melting Temperature (°C)	↓	↑	↓	↑
Stacking Energy (kcal/mol)	↓	↑	↓	↑
Enthalpy (kcal/mol)	↓	↑	↓	↑
Entropy (cal/mol/K)	↓	↑	↓	↑
Free Energy (kcal/mol)	↓	↑	↓	↑

Table 3: Overall comparison between Fusion sequences and Control

↑ indicates an overall increase

↓ indicates an overall decrease