Transcriptional regulation of multigene loci: multilevel control

NIALL DILLON AND FRANK GROSVELD

Recent studies indicate that different levels of control operate within multigene loci. In addition to regulatory sequences immediately flanking the genes, there are also elements that act over long distances on more than one gene. Competition for these elements among genes can influence both the level and timing of gene expression during development.

Many eukaryotic genes are organized into multigene loci in which genes with related function are coordinately expressed or differentially regulated during development. There are two key questions about the regulation of such loci. First, how is a large locus that may extend over hundreds of kilobases of DNA converted to a transcriptionally active state? Second, how are genes in such an activated region differentially regulated so that some are active and others silent at a particular developmental stage, and so that they can switch their expression state as development progresses?

Transcriptional control elements

In addressing the question of how loci are activated, we must consider the various elements that have been characterized on the basis of their ability to modulate the transcription of eukaryotic genes. It should be borne in mind that the terms used to describe these elements are functional definitions that entirely depend on the parameters of the assay used for the definition.

The best-characterized of the elements known to be important for gene regulation is the promoter: the region immediately upstream of each transcription initiation site. The promoter consists of binding sites for upstream factors and the basic transcription machinery, which is a very large complex with the TATA binding factor (TBP) as the central component. There are also other transcription factors, which act to promote or inhibit the basic machinery[1,2]. It has been shown that these elements are involved at many levels of control, such as tissue specificity, developmental and temporal specificity, and in regulating the level of gene expression. The presence of factors bound to an active promoter can usually be detected in chromatin as a region that is hypersensitive to DNase I digestion. It is important to note, however, that a promoter alone cannot generally drive efficient expression of a gene in cell transfection experiments and often fails to give any expression in transgenic mice.

A second type of element that can potentiate gene transcription is the enhancer. The first enhancers to be characterized were of viral origin, and were originally defined in transient transfection assays as fragments of DNA that would stimulate transcription of a gene *in cis* on the same plasmid[3,4]. The same functional assay was used to define cellular elements with similar properties, and it was found that these enhancers can show cell-type specificity that reflects the *in vivo* expression of their associated genes[5]. An enhancer typically contains a collection of sites that can bind activating or suppressing protein factors. Enhancers that function in transient assays can be created with relative ease by multimerization of individual protein-binding sites[6], suggesting that the functional definition may encompass a wide range of different elements. The term enhancer has since become used for almost any piece of DNA that stimulates transcription in almost any assay. However, it has also become clear that the presence of one or more enhancers is not sufficient to activate transcription in all test systems. In particular, in studies with transgenic mice almost all genes are not expressed properly despite the presence of enhancers. Expression levels are generally low or undetectable, and are not related to the copy-number of the transgene construct. The normal pattern of temporal and tissue-specific expression of the gene is also disturbed. Founder mice that have integrated the transgenic construct at a different chromosomal position often show different patterns of expression. This phenomenon, where the presence of other regulatory regions in the host genome at or near the site of integration of the construct influences the level and specificity of expression of the transgene, is called a position effect.

These observations led to the formulation of a new functional definition, that of the locus control region (LCR). The LCR was originally identified as a control region upstream from the human β-globin locus[7] (Fig. 1) and several other such elements have now been described that are important for the activation of various loci, including CD2 (Ref. 8), lysozyme[9] and the MHC (S. Carson and M. Wiles, pers. commun.). LCRs are characterized by tissue-specific, developmentally stable DNase-I-hypersensitive sites. Such sites are thought to be short stretches of DNA that are not complexed into nucleosomes.

An LCR has been functionally defined as an element that confers expression upon a transgene, to a level that is independent of its site of integration in the host genome but dependent on its copy number[7]. Thus LCR activity is quite differently defined from enhancer activity, using very different functional assays. In particular, the fact that the DNA is not integrated into chromatin in a transient assay allows the detection of a wide variety of sequences that have enhancer activity, but not LCR activity. The converse can also be the case, for example, in the β-globin locus the LCR hypersensitive sites HS3 and 4 (Fig. 1) do not act as enhancers in transient transfections[10,11], although they are powerful LCR elements[12–14]. HS2 shows activity in both types of assay[11,15,16], suggesting that there can be overlap between these two functional definitions. Given this overlap, it is not surprising that LCRs also contain multiple factor-binding sites, several of which have been shown to be important for activity of the element. However, in contrast to the ease with which enhancer activity can be obtained by binding site multimerization, we have as yet been unable to reproduce LCR activity using this approach.

This suggests that the precise arrangement of factor-binding sites may be critical for LCR function[17,18].

LCR function could conceivably be achieved by insulating the transgene from position effects. An insulating function of this type has been described for the *Drosophila* specialized chromatin structure (scs and scs') elements[19], sequences that were originally identified as DNase-I-hypersensitive sites. However, the possibility that the LCRs work by an insulating mechanism has been effectively excluded through an analysis of mice carrying a single copy of an LCR linked to a globin gene. These mice showed high levels of tissue-specific expression of the transgene[8,20], but low-level expression of the gene was detected in other tissues (D. Meijer, pers. commun.), indicating that it was not protected from positive position effects. Instead, extensive functional analysis has shown that the LCR is a dominant positive activator that may act by forming very stable complexes with genes, to the exclusion of other regulatory elements[12,16,18].

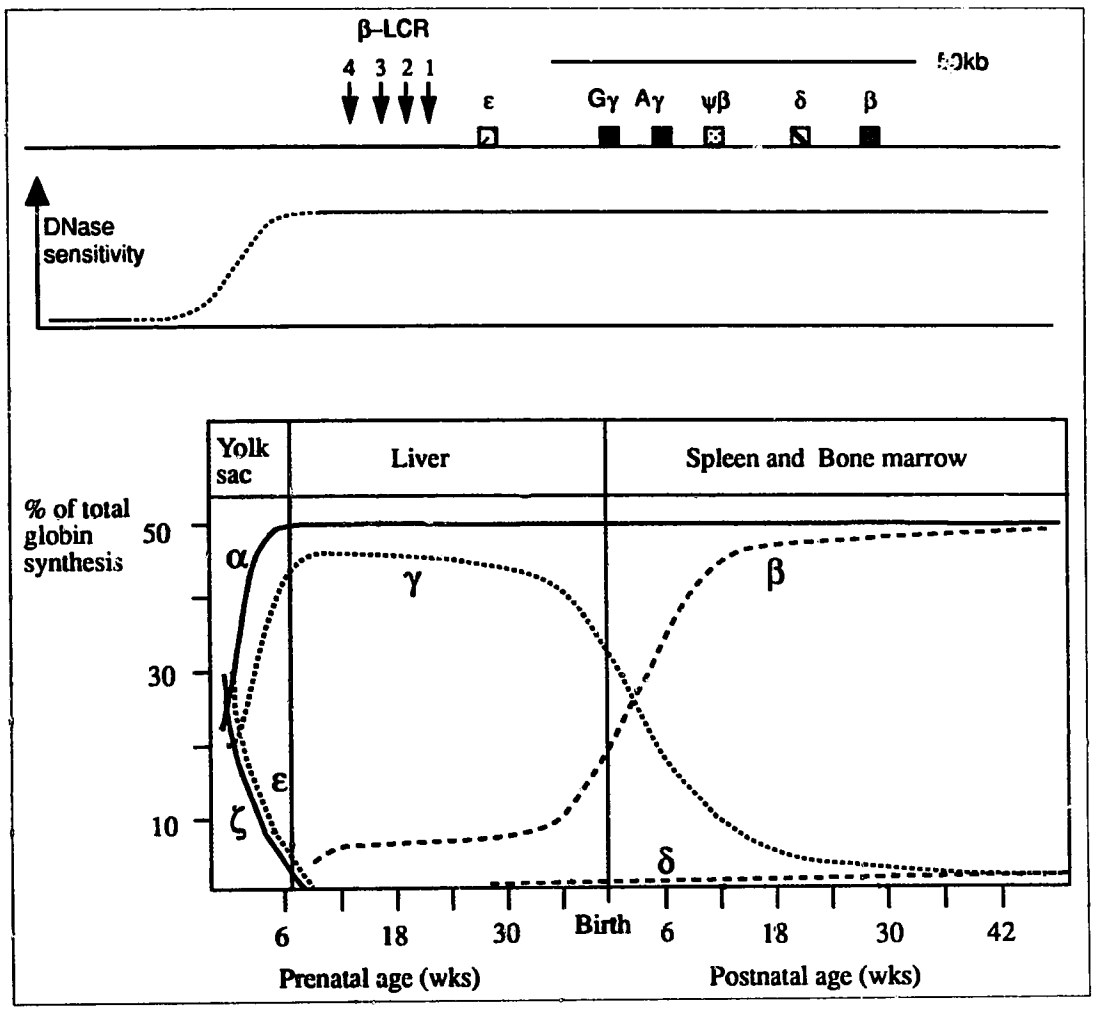

FIG 1

Top: Structure of the human β-globin locus. Vertical arrows indicate DNase I hypersensitive sites. DNase-I-sensitivity is shown graphically below the locus. The exact point of transition from sensitive to insensitive chromatin (dashed line) has not yet been determined.
Bottom: Production of human globin chains during development.

One property of the β-globin LCR is an ability to induce sensitivity to DNase I digestion over an extensive region of chromatin[21], suggesting that the locus may be present as a single large 'open' domain in erythroid cells. Although in the transgenic experiments the hypersensitive sites (HS) are restored[7], it is not known at present whether an equally large region of chromatin is affected, nor is it known whether the generation of HS is autonomous or requires the presence of other elements, such as a promoter.

In addition to promoters, enhancers and LCRs, other regulatory elements have been described. The *Drosophila* specialized chromatin structure elements mentioned above insulate regulatory elements from each other thereby preventing position effects, but they do not activate transcription nor are they matrix attachment sites[19,22]. In addition to the scs boundary elements, at least one boundary element has been identified in the *Drosophila* Bithorax complex that is essential for the correct temporal expression of the flanking genes[23]. No such elements have yet been functionally mapped in vertebrate loci, although they are widely assumed to be present. Sites that are thought to be nuclear matrix attachment regions[24] (MAR) or chromosomal scaffold attachment regions[25] (SAR) have not yet been shown to function as boundary elements in a transgenic experiment. This is also the case for the attachment sites (A elements) that are found at the boundaries of the DNase-I-sensitive chicken lysozyme gene domain[26]. These elements were part of a chicken lysozyme gene domain construct that had copy-number dependent expression when introduced into transgenic mice[27], but have not yet been shown to act as boundary elements when tested separately from the rest of the domain. The role of attachment sites is therefore as yet unclear. They may play a role in organizing chromosomal DNA for passage through the cell cycle. They might also have a part in nonspecific repression mediated by higher order chromatin structures. General functions such as these are likely to be antagonistic to the achievement of specific programmes of gene activation during development, and the role of an LCR may well be to override such effects and convert a locus to a stable transcriptionally active site.

Gene regulation in multigene loci

Detailed analysis of the mechanisms of regulation in multigene loci is now becoming possible with the development of methods for generating transgenic mice carrying large DNA fragments. The human β-globin cluster has been particularly useful for these types of studies, as all of the sequences required for full expression of the locus in chromatin have been defined, and the LCR is located on one piece of DNA. The globin genes are arranged in the order in which they are expressed during development (Fig. 1) with

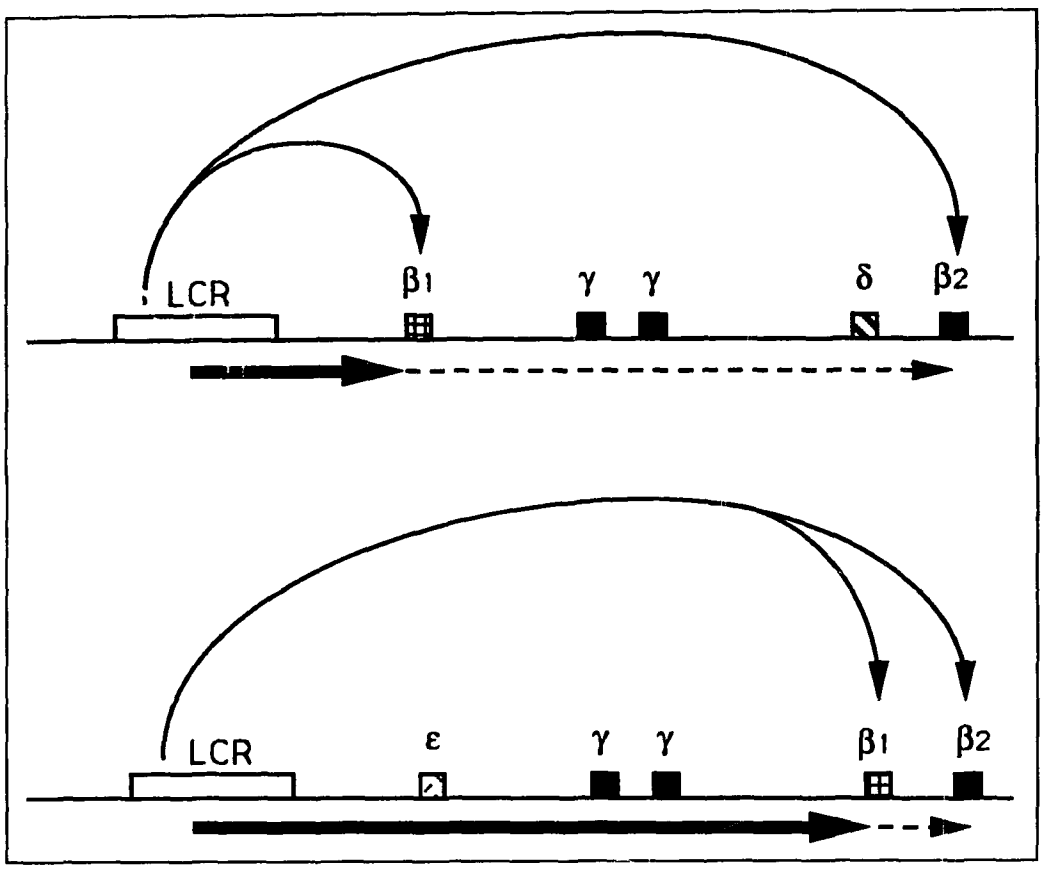

FIG 2

Scheme for differentiating between tracking and frequency models by examining the effect of varying the relative distance of two β-globin genes from the LCR. Loop formation is indicated by curved arrows. A straight solid arrow indicates a tracking effect. A straight dashed arrow indicates tracking blocked or attenuated by the presence of an active gene.

ε-globin showing an embryonic and early foetal profile and γ-globin being expressed in the late embryonic stage and throughout the foetal period[28]. Expression of β-globin increases during the foetal period and this becomes the predominant species in the adult (reviewed in Ref. 29). The LCR is required for activation of all of the genes, since in naturally occurring mutants in which the LCR is deleted, the genes are intact, but silent[30,31].

Extensive studies in transgenic mice have indicated that a major part of the developmental specificity of the β-globin genes resides in the promoters[20,32,33]. For example, it is clear that the promoters of the γ-globin genes bind one or more repressing factors which silence transcription in the adult, and that this process can be reversed by mutations in the binding sites for these factors[34,35]. A substantial contribution is also made by differential activity of individual hypersensitive LCR sites during development[36]. A further level of regulation appears to result from competition among genes for activation by the LCR. This type of intergenic competition was originally documented in transient expression assays with the SV40 enhancer[37,38] and, more recently, with the chicken β-globin enhancer[39]. The latter study showed that active transcription is required for this process.

Genetic evidence for competition in the β-globin locus *in vivo* comes from the observation that in individuals with hereditary persistence of foetal haemoglobin, increased adult expression of a γ-globin gene caused by mutations in the gene promoter (reviewed in Ref. 40) results in a corresponding down-regulation of the β-globin gene from the same allele[41]. Mutations and small deletions that inactivate the β-globin gene promoter cause only a slight increase in γ expression in the adult[40], suggesting that competition in the locus is polar. This model first arose from transgenic mouse studies[42] which showed that while a gene proximal to the LCR could compete the expression of a distal gene, one located distally was unable to compete effectively. Polar competition is interesting mechanistically and provides us with a potential means to investigate how the LCR acts over long distances. A variety of mechanisms have been proposed to explain the action of genetic control elements over long distances. These include polymerase tracking, the spreading of chromatin structures, and direct contact between separated elements through loop formation. Although in recent years the loop hypothesis has been supported by transient expression[43] and *in vitro* studies[44], it should be remembered that, so far, the only direct evidence for contact between two separated control elements as a mechanism for transcriptional activation in chromatin is the phenomenon of transvection in *Drosophila*[45].

Polar competition could be explained by either tracking or spreading models. In these models, proteins (e.g. polymerases) or protein complexes move along the DNA in a linear fashion, altering the chromatin structure in the process. The blocking or attenuation of such complexes by the first gene encountered would result in a polar effect. Alternatively, polarity might be explained by a frequency model, in which movement of DNA in solution leads to random contact between promoter and LCR and the formation of stable loops. According to this stochastic model, the proximal γ-globin genes would interact with the LCR more frequently and this would accentuate the effect of a stronger promoter during the foetal stage. The less frequent interaction of the distally located β-globin gene with the LCR would place it at a disadvantage even when its promoter is stronger in the adult stage, and would only allow it to be expressed fully when the γ-globin genes are completely silenced by stage-specific factors. These models could be directly tested in mice transgenic for the complete β-globin locus[46], by placing a second β-globin gene at different points in the locus (Fig. 2). A tracking model predicts that the competitive effect of the proximal β-globin gene will be similar wherever it is located. A frequency model predicts that the proximal β-globin gene will compete less effectively as it is moved away from the LCR and towards the distal β-globin gene (Fig. 2).

Implications for other multigene loci

From the study of the β-globin locus, it has become clear that individual regulatory elements in a multigene locus are not necessarily restricted to acting on any one gene. Instead there can be an extra level of complexity with elements acting over long distances on several genes. The result of this is a situation where expression of one gene can affect the timing and the level of expression of another gene. Regulation of other multigene loci probably involves a similar type of control. For example, the order of genes within the *Hox* clusters has been maintained from *Drosophila* to humans, suggesting that this regulatory principle may also operate in these loci[47]. Although it has been shown that the intergenic flanking regions of certain Hox genes can confer correct temporal expression in transgenic mice[48,49], it is not yet known whether the genes are also expressed at the correct level. Both the timing and the level of expression of each gene are likely to be important for pattern formation. Since correct temporal expression of a given gene requires regulatory sequences immediately flanking the next gene, it is probable that its level of expression will depend on the presence or absence of neighbouring genes.

Advances in investigating the control of multigene loci will continue to depend on stringent functional analysis *in vivo*. The ability to introduce large DNA molecules into mice will be important for such studies. Homologous recombination also offers a particularly good opportunity to carry out this type of analysis, since any putative regulatory region can be deleted, modified or inserted *in situ*. It is likely that there will be significant progress in understanding the regulation of complex gene loci in the next few years.

Acknowledgements

We thank M. Antoniou for helpful comments on the manuscript and all our colleagues for useful discussion. Our work is supported by the MRC (UK).

References

1 Dynan, W. (1989) *Cell* 58, 1–4
2 Sharp, P.A. (1992) *Cell* 68, 819–821
3 Banerji, J., Rusconi, S. and Schaffner, W. (1981) *Cell* 27, 299–308
4 Moreau, P. *et al.* (1981) *Nucleic Acids Res.* 9, 6047–6068
5 Picard, D. (1985) in *Oxford Surveys on Eukaryotic Genes* (Vol. 2), pp. 24–48, Oxford University Press
6 Zenke, M. *et al.* (1986) *EMBO J.* 5, 387–397
7 Grosveld, F., Blom van Assendelft, G., Greaves, D. and Kollias, G. (1987) *Cell* 51, 975–985
8 Greaves, D.R., Wilson, F.D., Lang, G. and Kioussis, D. (1989) *Cell* 56, 979–986
9 Sippel, A. *et al.* (1992) in *Transgenic Animals* (Grosveld, F. and Kollias, G., eds), pp. 1–26, Academic Press
10 Hug, B., Moon, A. and Ley, T. (1992) *Nucleic Acids Res.* 20, 5771–5778
11 Tuan, D.Y., Solomon, W.B., London, I.M. and Lee, D. (1989) *Proc. Natl Acad. Sci. USA* 86, 2554–2558
12 Pruzina, S. *et al.* (1991) *Nucleic Acids Res.* 19, 1413–1419
13 Philipsen, S., Talbot, D., Fraser, P. and Grosveld, F. (1990) *EMBO J.* 9, 2159–2167
14 Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) *Nucleic Acids Res.* 18, 3503–3508
15 Sorrentino, B., Ney, P., Bodine, D. and Nienhuis, A.W. (1990) *Nucleic Acids Res.* 18, 2721–2731
16 Talbot, D. and Grosveld, F. (1991) *EMBO J.* 10, 1391–1398
17 Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) *EMBO J.* 9, 2169–2177
18 Philipsen, S., Pruzina, S. and Grosveld, F. *EMBO J.* (in press)
19 Kellum, R. and Schedl, P. (1991) *Cell* 64, 941–950
20 Dillon, N. and Grosveld, F. (1991) *Nature* 350, 252–254
21 Forrester, W.C. *et al.* (1990) *Genes Dev.* 4, 1637–1649
22 Kellum, R. and Schedl, P. (1992) *Mol. Cell. Biol.* 12, 2424–2431
23 Gyurkovics, H., Gausz, J., Kummer, J. and Karch, F. (1990) *EMBO J.* 9, 2579–2585
24 Cockerhill, P. and Garrard, W. (1986) *Cell* 44, 273–282
25 Gasser, S.M. and Laemmli, U.K. (1986) *Cell* 46, 521–530
26 Stief, A., Winter, D.M., Stratling, W.H. and Sippel, A.E. (1989) *Nature* 341, 343–345
27 Bonifer, C., Vidal, M., Grosveld, F. and Sippel, A. (1990) *EMBO J.* 9, 2843–2848
28 Peschle, C. *et al.* (1985) *Nature* 313, 235–238
29 Stamatoyannopoulos, G. and Nienhuis, A. (1987) in *The Molecular Basis of Blood Diseases* (Stamatoyannopoulos, G., Nienhuis, A., Leder, P. and Majerus, P., eds), pp. 66–105
30 Kioussis, D. *et al.* (1983) *Nature* 306, 662–666
31 Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989) *Proc. Natl Acad. Sci. USA* 86, 7470–7474
32 Shih, D.M., Wall, R.J. and Shapiro, S.G. (1990) *Nucleic Acids Res.* 18, 5465–5472
33 Raich, N. *et al.* (1990) *Science* 250, 1147–1149
34 Berry, M., Grosveld, F. and Dillon, N. (1992) *Nature* 358, 499–502
35 Mantovani, R. *et al.* (1988) *Nucleic Acids Res.* 16, 4299–4313
36 Fraser, P., Pruzina, S., Antoniou, M. and Grosveld, F. (1993) *Genes Dev.* 7, 106–113
37 deVilliers, J., Olson, C., Banerji, J. and Schaffner, W. (1982) *Cold Spring Harbor Symp. Quant. Biol.* 47, 911–919
38 Wasylyk, B., Wasylyk, C., Augerean, P. and Chambon, P. (1983) *Cell* 32, 503–514
39 Choi, O.R. and Engel, J.D. (1988) *Cell* 55, 17–26
40 Poncz, M., Henthorn, P., Stoeckert, C. and Surrey, S. (1988) in *Oxford Surveys on Eukaryotic Genes* (Vol. 5), pp. 163–203, Oxford University Press
41 Giglioni, B. *et al.* (1984) *EMBO J.* 11, 2641–2645
42 Hanscombe, O. *et al.* (1991) *Genes Dev.* 5, 1387–1394
43 Muller, H., Soso, J. and Schaffner, W. (1989) *Cell* 58, 767–777
44 Su, W., Jackson, S., Tjian, R. and Echols, H. (1991) *Genes Dev.* 5, 820–826
45 Tartof, K. and Henikoff, S. (1991) *Cell* 65, 201–203
46 Strouboulis, J., Dillon, N. and Grosveld, F. (1992) *Genes Dev.* 6, 1857–1864
47 Krumlauf, R. (1993) *Trends Genet.* 9, 106–112
48 Puschel, A., Balling, R. and Gruss, P. (1991) *Development* 112, 279–287
49 Whiting, J. *et al.* (1991) *Genes Dev.* 5, 2048–2059

N. DILLON AND F. GROSVELD ARE IN THE LABORATORY OF GENE STRUCTURE AND EXPRESSION, MRC NATIONAL INSTITUTE FOR MEDICAL RESEARCH, THE RIDGEWAY, MILL HILL, LONDON, UK NW7