

Using Interval Logic for Order Assembly

Zhan Cui

Advanced Computation Laboratory
Imperial Cancer Research Fund
61 Lincoln's Inn Fields
LONDON WC2A 3PX
United Kingdom

email: cui@acl.lif.icnet.uk

Abstract

Temporal logic, in particular, interval logic has been used to represent genome maps and to assist genome map constructions. However, interval logic itself appears to be limited in its expressive power because genome mapping requires various information such as partial order, distance and local orientation. In this paper, we first propose an integrated formalism based on a spatial-temporal logic where the concepts of metric information, local orientation and uncertainty are merged. Then, we present and discuss a deductive and object-oriented data model based on this formalism for a genetic deductive database, and the inference rules required. The formalism supports the maintenance of coarser knowledge of unordered, partially ordered and completely ordered genetic data in a relational hierarchy. We believe that this integrated formalism also provides a formal basis for designing a declarative query language.

Introduction

The proliferation of data from the Human Genome Project has brought new challenges for researchers in the database research community. At the Imperial Cancer Research Fund, we are investigating how to build a genetics database in the context of deductive and object-oriented databases. More specifically, our system will be developed on the IDEA database system currently developed within the ESPRIT EP6333 project¹. The focus of this project is the design and support of a high-level conceptual interface for a knowledge processing environment, based on an object-oriented conceptual model, a declarative rule language, and a procedural update language. The resulting system will provide advanced database technologies to raise the quality of interaction with users and applications as well as the quality and efficiency of data management.

One of our objectives in using IDEA technology is to provide an efficient support for storing genome map

data and for map construction. Map construction is the use of experimental data about known loci from physical or genetic measurements to derive their relative order and separation on a map. Interpretations of raw data provide partial orders and distances in local orientations which must be combined into larger maps to describe whole chromosomes. Biological laws and experimental techniques provide rules for inferring order and distance from single experiments and for the integration of data from different sources. The challenge of map construction is to reconcile and integrate local information into a global map and to update this continually. There is an urgent need to develop a well founded formalism to support these activities in an 'intelligent' knowledge-based system.

Map objects such as clones can be considered as intervals of Allen's temporal logic (Allen 1983, 1984). Work on using interval logic has been reported and surveyed in Letovsky and Berlyn (1992), Graves (1993), Guidi and Roderick (1993). However, we found most of these systems failed to incorporate the necessary information needed in order assembly such as partial order, metric information, orientation, uncertainty and ambiguity, which have been regarded as essential (also Guidi and Roderick 1993). In addition, the lack of a formalism imposes great difficulty in developing a good query language, especially a declarative query language. This situation could be improved by integrating and extending the formalisms developed in the spatial and temporal logic community. In this paper, we will discuss the problems which need to be addressed by such a formalism and then propose a formalism where the concepts of metric information, local orientation and uncertainty are merged.

The remainder of this paper is structured as follows. In the next section, we discuss the basic requirements for order assembly and problems that need to be addressed in an 'intelligent' system. Then we present the proposed formalism and its various extensions. A set of inference rules based on the proposed formalism is also described in this section. Following the discussion on the data model for molecular biology based on the proposed formalism, we discuss the implemen-

¹IDEA stands for Intelligent Database Environment for Advanced Applications. This paper reflects the opinions of the author and not necessarily those of the consortium.

tation and its related issues. Finally, we conclude the paper with the main features of this formalism and the further work.

Problems

Genome maps fall into several basic categories: genetic linkage maps, cytogenetic maps and molecular-based physical maps. Figure 1 from Honda, Parrot, Smith and Lawrence (1993) illustrates some typical genome maps at different resolutions of details from the lowest level of chromosome bands to the highest level of nucleotide sequences.

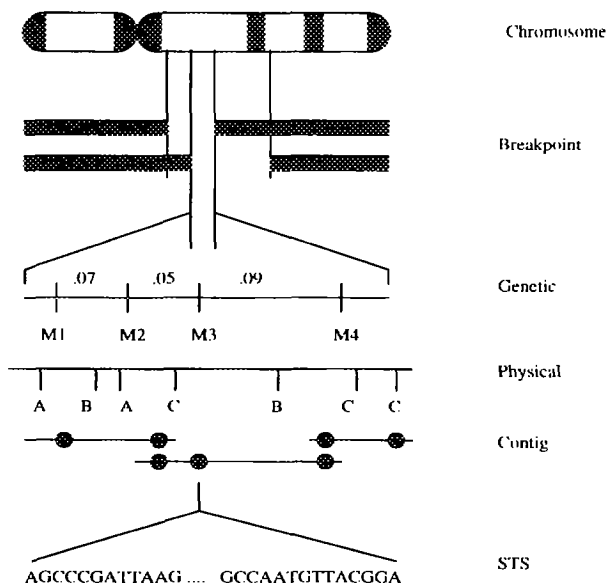


Figure 1: Genome related maps

As can be seen from the figure, these maps are basically 1-dimensional and are akin to problems dealt with by temporal logic. Indeed, several systems based on temporal logics have been built. Letovsky and Berlyn (1992) described a point set ontology for their constraint-based system. In their system, only one relation *before* is used, and local ordering windows are introduced to handle the orientation uncertainty. An interesting set of inference rules are presented to derive the global ordering. Guidi and Roderick (1993) planned to use Allen's interval-like logic to represent partial ordering to cope with uncertainty in the fragments of DNA sequences, but from their short survey paper it is not clear how they propose to use them. Honda et al. (1993) proposes an object-oriented data model which is also based on the general idea of linear ordering. One slightly different approach is reported in Graves (1993), who developed a general knowledge-based system based on connection graph and a single order relation.

There are, however, some significant differences between temporal logic and genomics. We can view both

DNA fragments and probes as intervals, but their orders are usually only true in some local frame of reference. Secondly, experiments are designed to produce overlapping fragments and we normally do not know the precise beginnings and endings of intervals. The partial orders from experiments are often uncertain due to experimental errors. Thirdly, conflicting orders may be reported by different experiments, which must be represented and stored in the database because those conflicting results might have some biological explanations yet to be discovered. Furthermore, the algorithms for constructing orders of genetic markers use various sources of information such as partial ordering, distance and orientation. All this demands a formalism more expressive than Allen's interval logic: it must be flexible enough to support the diversity of raw data required by different map construction algorithms. Some of the key requirements are:

Imprecise data: Imprecise information abounds in genetics. Allen's interval logic imposes precise information between intervals. Although uncertainty relations can be represented by disjunctions, the computational cost is high. In recent papers, Freksa (1992) has proposed "conceptual neighbourhood" to encode imprecise information. It would be very appealing to develop a formalism along this line where imprecise information can be represented in a refinement hierarchy.

Metric data: Metric information must be represented. Geneticists often use it to derive new order information. Several successful mapping systems require distances between genetic markers.

Local orientation: This is missing from most of the spatial and temporal logics, but must be supported. Due to the nature of the experiment, the marker ordering information obtained is usually local to a particular reference frame.

Relations between DNA fragments: Most of the systems we are aware of support only a very limited number of relations when in fact there are many possibilities. The relations between two DNA fragments reported from experiments can be described as separated, adjacent, overlapping, contained, etc. A set of binary relations is needed to represent the coarseness of our knowledge about two fragments or intervals. Furthermore, we also need to represent data which may be unordered, partially ordered or completely ordered.

Uncertainty: Uncertainty can come from different sources in this context. Experimental data are usually associated with experimental errors and ambiguous results may be reported by different experiments. Even if we have certain data, uncertainty may be present after one cycle of inferences (Freksa 1992). Uncertainty management can keep a database consistent and report inconsistency. In addition, it may provide a way to manage inconsistency in the database.

Inference rules: A set of new inference rules is needed to reason with the partial order, local orientation and ambiguity, etc in this domain. Letovsky and

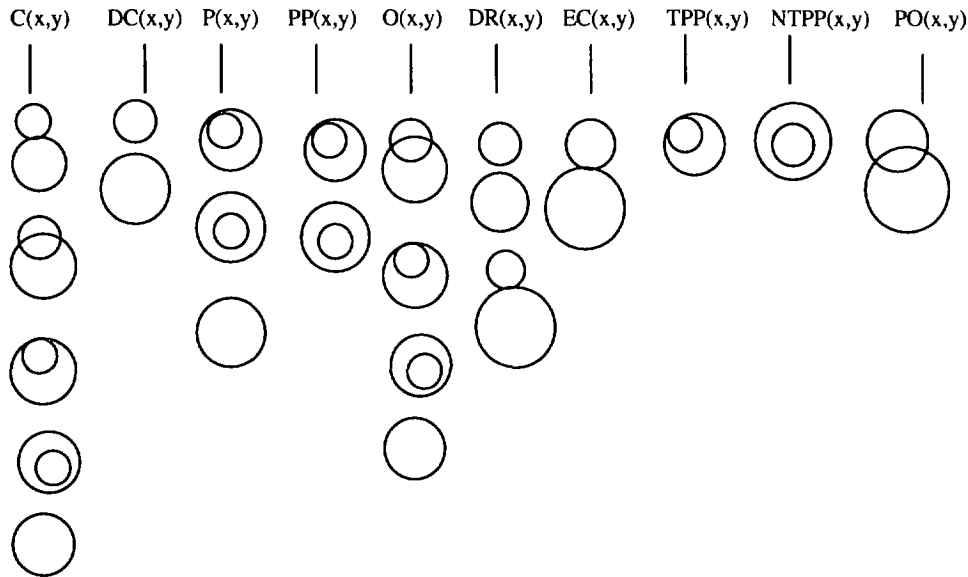


Figure 2: A set of sample configurations (in 2D) modelling the defined relations.

Berlyn (1992) have shown the usefulness of these rules in their CPRON program.

Declarative query language: The importance of declarative query language for a genetic database has been evidenced by the use of Prolog in several genetic databases (Kazic et al. 1990, Yoshida et al. 1992). A good query language for a genome map database will give the user great flexibility to construct complex queries including recursive queries.

A Proposed Formalism

In a genome database, especially a database for genome mapping information, there is a need to represent different levels of information from cytogenetic through linkage map to physical and sequence data. Although Allen's interval logic provides a set of base relations, no higher level relations are defined. Recent work of Freksa (1992) has shown how weaker spatial/temporal relations can be defined using semi-intervals, but it is not clear how to arrange them in a relational hierarchy. The work of Randell et al. (1992), Cui et al. (1993) on spatial logic has presented their binary relations in a relational hierarchy which can be exploited in several ways. This logic can also be extended to define all of the 13 mutually exclusive interval relations of Allen's logic. In this section, we will start from this formalism, and then move onto the extensions to handle the problems mentioned above.

The Formalism

The basic ontological entity in this formalism is a region; note that boundaries, lines and points are not regions. Regions are non empty. Regions in the theory support either a spatial or temporal interpretation.

Informally, these regions may be thought to be potentially infinite in number, and any degree of connection between them is allowed in the intended model, from external contact to identity in terms of mutually shared parts. The formalism supports two or three dimensional interpretations (or higher dimensions!). It is based upon Clarke's (1981, 1985) calculus of individuals based on "connection" and is expressed in the many sorted logic LLAMA (Cohn 1987).

We have used the standard conventions of first order logic. These are summarised in Table 1. Note that the axioms in the following would be more complex without using many sorted logic. For example, the first axiom: $\forall x C(x, x)$ would become $\forall x [\text{Region}(x) \rightarrow C(x, x)]$.

The basic part of the formalism assumes one primitive dyadic relation: $C(x, y)$ read as 'x connects with y'. The relation $C(x, y)$ is reflexive and symmetric, i.e. x connects with x and if x connects with y , then y connects with x . We can give a topological model to interpret the theory, namely that $C(x, y)$ holds when the topological closures of regions x and y share a common point. Two axioms are thus introduced.

- (1) $\forall x C(x, x)$
- (2) $\forall xy [C(x, y) \rightarrow C(y, x)]$

Using $C(x, y)$, a basic set of dyadic relations are defined: 'DC(x, y)' ('x is disconnected from y'), 'P(x, y)' ('x is a part of y'), 'PP(x, y)' ('x is a proper part of y'), ' $x = y$ ' ('x is identical with y'), 'O(x, y)' ('x overlaps y'), 'DR(x, y)' ('x is discrete from y') 'PO(x, y)' ('x partially overlaps y'), 'EC(x, y)' ('x is externally connected with y'), 'TPP(x, y)' ('x is a tangential

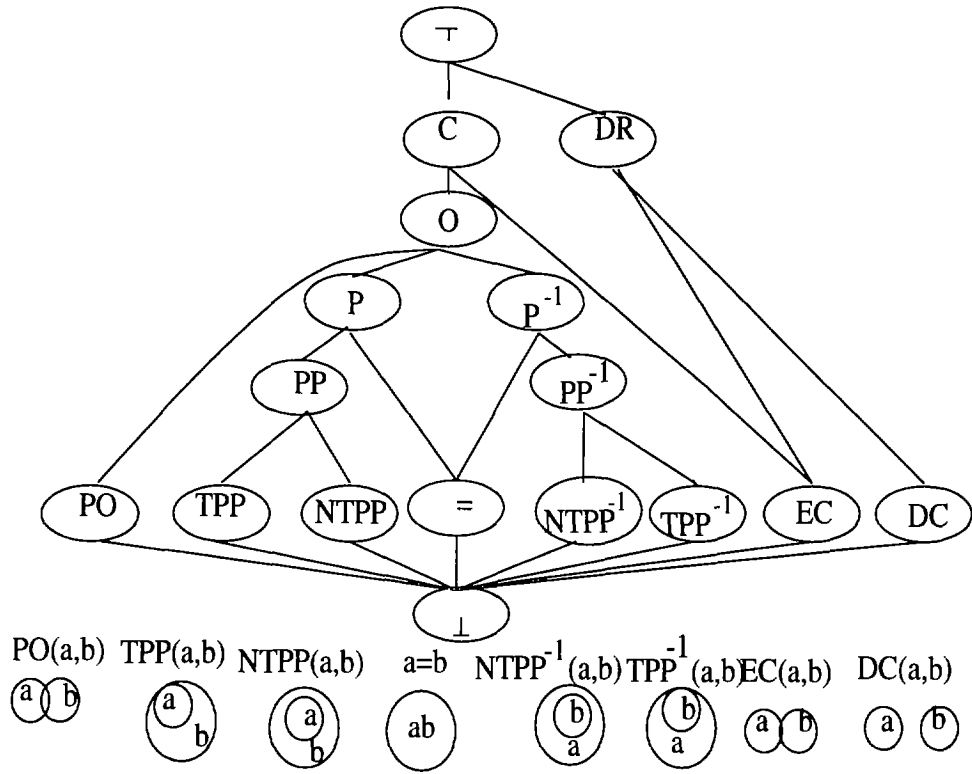


Figure 3: A lattice defining the subsumption hierarchy of the dyadic relations defined solely in terms of the primitive relation $C(x, y)$.

negation:	\neg	not	pairwise disjoint.
conjunction:	\wedge	and	(3) $DC(x, y) \equiv_{def} \neg C(x, y)$
disjunction:	\vee	or	(4) $P(x, y) \equiv_{def} \forall z[C(z, x) \rightarrow C(z, y)]$
implication:	\rightarrow	if then	(5) $PP(x, y) \equiv_{def} P(x, y) \wedge \neg P(y, x)$
bi-implication:	\leftrightarrow	if and only if	(6) $x = y \equiv_{def} P(x, y) \wedge P(y, x)$
existential quantifier:	\exists	there exists	(7) $O(x, y) \equiv_{def} \exists z[P(z, x) \wedge P(z, y)]$
universal quantifier:	\forall	for all	(8) $PO(x, y) \equiv_{def} O(x, y) \wedge \neg P(x, y) \wedge \neg P(y, x)$
definition:	\equiv_{def}	equivalent	(9) $DR(x, y) \equiv_{def} \neg O(x, y)$
			(10) $EC(x, y) \equiv_{def} C(x, y) \wedge \neg O(x, y)$
			(11) $TPP(x, y) \equiv_{def} PP(x, y) \wedge \exists z[EC(z, x) \wedge EC(z, y)]$
			(12) $NTTPP(x, y) \equiv_{def} PP(x, y) \wedge \neg \exists z[EC(z, x) \wedge EC(z, y)]$
			(13) $P^{-1}(x, y) \equiv_{def} P(y, x)$
			(14) $PP^{-1}(x, y) \equiv_{def} PP(y, x)$
			(15) $TPP^{-1}(x, y) \equiv_{def} TPP(y, x)$
			(16) $NTTPP^{-1}(x, y) \equiv_{def} NTTPP(y, x)$

Table 1: Symbols used in the axiomatization

proper part of y) and 'NTTPP(x, y)' (' x is a nontangential proper part of y '). The relations: P, PP, TPP and NTTPP being non-symmetrical support inverses: P^{-1} , PP^{-1} , TPP^{-1} and $NTTPP^{-1}$. Of the defined relations, DC, EC, PO, =, TPP, NTTPP and the inverses for TPP and NTTPP are provably mutually exhaustive and

A pictorial representation of the relations defined above is given in Figure 2 and their relational hierarchy in Figure 3.

The formalism has a temporal interpretation by introducing another primitive relation $B(x, y)$ - of x being before y . The formal definitions for Allen's interval relation before ($<$), meets (m), overlaps (o), during

(d), starts (s), finishes (f) and their inverses (>, mi, oi, di, si and fi) are given below

$\langle (x, y) \equiv_{def} DC(x, y) \wedge B(x, y)$
 $\rangle (x, y) \equiv_{def} DC(x, y) \wedge B(y, x)$
 $m(x, y) \equiv_{def} EC(x, y) \wedge B(x, y)$
 $mi(x, y) \equiv_{def} EC(x, y) \wedge B(y, x)$
 $o(x, y) \equiv_{def} PO(x, y) \wedge \exists z[m(z, x) \wedge \neg m(z, y)]$
 $oi(x, y) \equiv_{def} PO(x, y) \wedge \exists z[mi(z, x) \wedge \neg mi(z, y)]$
 $d(x, y) \equiv_{def} NTPP(x, y)$
 $di(x, y) \equiv_{def} NTPP(y, x)$
 $s(x, y) \equiv_{def} TPP(x, y) \wedge \exists z[m(z, x) \wedge m(z, y)]$
 $si(x, y) \equiv_{def} TPP(x, y) \wedge \exists z[mi(z, x) \wedge mi(z, y)]$
 $f(x, y) \equiv_{def} DC(x, y) \wedge \exists z[mi(z, x) \wedge mi(z, y)]$
 $fi(x, y) \equiv_{def} DC(x, y) \wedge \exists z[m(z, x) \wedge m(z, y)]$

Now the leaf nodes in Figure 3 can be further refined to include all of the Allen's 13 interval relations just defined when used to represent one dimensional space.

Metric extension

So far we have principally concentrated on developing a purely qualitative calculus. However, this can never be a replacement for metric information, but rather should complement a metric representation. Indeed, distance is considered to be an important source of information in genome mapping algorithms (Letovsky and Berlyn 1992). In CPRP, distance is used to form several inference rules and Guidi and Roderick (1993) show how distances alone can sometimes give precise orders.

There is a close relationship between the metric information and the binary relations described in section 3.1 and distances when only one dimensional space is considered. All of the 13 base relations can be defined by using the relations between the endpoints of intervals (Allen 1983, Kauts and Ladkin 1991, Freksa 1992). The correspondence between them has been shown in Table 2. However, we should note that endpoints are qualitative in nature because the precise measure of endpoints, particularly of clones, may not be available.

We should also note that the distance between intervals can be used as a measure of their endpoints. Hence, all the techniques in dealing with numbers can be readily used together with the techniques for interval reasoning. This provides a natural link between topological relations and distances. We can use them to develop a fast inference rule for the derivation of new order relations. The techniques presented in qualitative simulation and spatial reasoning in databases (Cui et al. 1992, 1993) should also be useful here.

Orientation

Due to the nature of experiments, locus order is usually only reported with reference to a local orientation and thus with an uncertainty over the global order. Letovsky and Berlyn (1992) have presented an

a before b	$e(a) < s(b)$
a = b	$s(a) = s(b)$ $e(a) = e(b)$
a meets b	$e(a) = s(b)$
a overlaps b	$s(a) > s(b)$ $e(a) < e(b)$ $e(a) > s(b)$
a during b	$s(a) > s(b)$ $e(a) < e(b)$
a starts b	$s(a) = s(b)$ $e(a) < e(b)$
a finishes b	$s(a) > s(b)$ $e(a) = e(b)$

Table 2: Correspondence between interval relations and their endpoints. $s(x)$ denotes the beginning of interval x and $e(x)$ the ending.

excellent treatment of orientation uncertainty by introducing local orientation windows. This idea of local orientation can be easily incorporated into the formalism just presented since the basic formalism does not require orientation to be specified. Hence, a special inference layer can be introduced to handle local orientation. As we associate a set of genetic markers with a particular orientation without reference to the global orientation, the derived facts within that particular set of markers will still hold even if we move them around. This is particularly useful in the genetic domain since the exact position of a locus may not be fixed due to incomplete knowledge.

The techniques for merging two local orientations (LO) use the facts that only one base relation can hold for a pair of markers. We can emphasise that markers occupy space by calling them intervals. The base relation (order relationship) between two intervals in two separate local orientations can not be inconsistent. Thus we can deduce the relative order of two local orientations and merge their data. For example, we can conclude that LO_j is the inverse of LO_i from

A before B before C, in LO_i

D before C before B, in LO_j

Therefore,

A before B before C before D, in LO_i

We have constructed a complete composition table for combining two local orientations for all the combinations of the base relations. Not all the base relations

give orientation information. For example, TPP and NTPP can be true in any orientation frame. The base relations that do fix orientation information are before, after, meet and its inverse, overlap and its inverse. The composition table not only provide an efficient way to merge two local orientations, but also can be used to detect inconsistencies in the raw data. For example, unlike the merging of local orientations above, information given in two local orientations can lead to contradiction no matter how we interpret the two local orientations and further experimentation may be required.

Inference Rules

The formalism supports a new set of inference rules. By exploiting the lattice structures presented in the formalism, we may be able to implement these inference rules effectively. For example, given two relations: $R_1(A, B)$ and $R_2(B, C)$, the relation between A and C can be derived. This allows us to calculate a composition table (or transitivity table) for all combinations of the base relations. There are two composition tables in our formalism: one for the 13 Allen's interval relations and one for the base relations (before the temporal extension) presented in Figure 3. Due to the space limitation, we could not present the two tables here. However, they can be found in Allen (1984) and Randell et al. (1992). We have also developed other inference rules similar to CPRP's (Letovsky and Berlyn 1992). In CPRP, the merging of local orientation windows was considered as an application of inference rules. In our case, we regard this as preprocessing because we found that to do otherwise would unnecessarily complicate our inference engine.

By expressing all the possible relations between intervals, our formalism allows maximal use of available information in map assembly. For example, the relation 'contains' is usually considered to be no use in the genome mapping processing. However, this information can be used to derive orders in our representation. Figure 4 shows an example of the use of interval logic in automated reasoning. Suppose we have data from three separate experiments (1, 2 and 3) concerning three loci: A , B and C . Experiment 1 reveals a partial overlap between A and C : $PO(A, C)$, which may have the two local orientations as shown. Experiment 2 shows that locus C lies before B : $B(C, B)$, and Experiment 3 detects that locus A contains locus B , expressed as B is a proper part of A : $PP(B, A)$. Without using $PP(B, A)$, there are two possible cases shown in Figure 4. However, this ambiguity can be resolved by $PP(B, A)$ because case **b** then becomes inconsistent with the data.

Data Model

The integrated formalism has many properties which could not be readily and efficiently represented and exploited by conventional databases. For example, Fig-

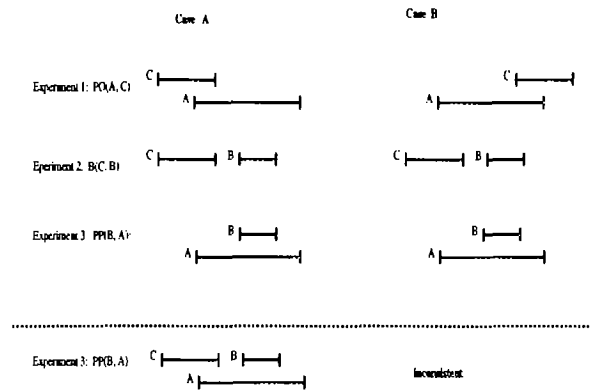


Figure 4: An example of containment relations used in order reasoning

ure 3 suggests that the binary relations of DNA fragments can be represented in an inheritance hierarchy, which could not be efficiently supported in a relational data model. In the following, our discussion is based on a deductive and object-oriented representation.

Object-oriented modelling provides a natural means of representing the complex, often composite structures in molecular biology by the inheritance of attributes and behaviours. The deductive capabilities enable us to store biological rules in the database in the form of Prolog (Kazic et al. 1990). Although databases based on object-oriented models have been successfully used in several human genome projects (Yoshida et al. 1992, Honda et al. 1993), the work on deductive and object-oriented data models (DOOD) is less advanced. One of the difficulties is due to the lack of a suitable formalism to encode the biological knowledge in DOOD. We believe that the proposed formalism may provide a basis for representing this body of knowledge.

In our data model, we have identified the following classes, each of which corresponds to independent entities in the domain: chromosome, probe, clone, gene, YAC, cosmid, STS fragments, protein, enzyme, restriction sites, etc.; some other entities are related to experiments: agents, protocols, etc, and still other entities correspond to binary relations such as orders. Most of these entities can be straightforwardly defined as classes with simple attributes and simple passive rules (Ceri and Manthey 1993). For example,

```
class chromosome_band =
  attributes
    name: string,
    chromosome: int,
    arm: string,
    length: real,
    first_locus: locus,
    last_locus: locus
end
```

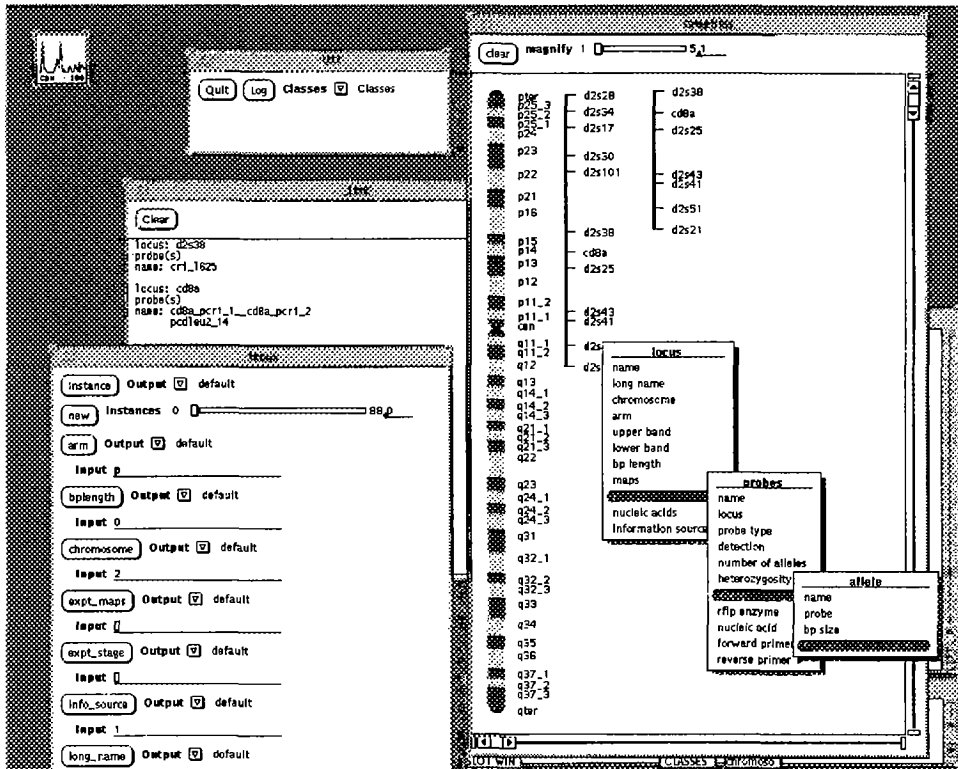


Figure 5: A screendump of our graphical interface for our prototype implementation

The binary relationships between two DNA fragments are also defined as objects. However, these objects are classified according to the hierarchy in Figure 3 with the leaf nodes further split to include all 13 of Allen's interval relations. More specifically, we define one class for each node in Figure 3. The top level class (or node) includes the generic attributes such as map-objects², distances and local orientations. The rules for translating metric information to topological one are defined as passive rules (also see Ceri and Manthey 1993, Jonker et al. 1994). The leaf classes represent the precise relations between two genetic fragments we know about in the database. Imprecise information is captured by the higher level classes with the highest level class representing completely unordered fragments. The ambiguous orders are represented by multiple instances in our database.

Implementation

The data model described in the previous section can be implemented on a deductive and object-oriented database. As we are participating the IDEA project, we have implemented our system on the IDEA prototype deductive and object-oriented database. How-

²Map objects (equivalent in function to intervals in Allen's logic) are object identifiers for DNA fragments, ordered maps, etc.

ever, the prototype has not yet provided all the functionality required at this stage. Instead, we have implemented our static data model and some passive rules in the IDEA prototype database, and implemented the inference mechanism and active behaviours in SLOT (Cui, Fox and Hearne 1993, Hearne et al. 1994, Cui and Fox 1994) – a specification language for building deductive and object-oriented theories, which combines the advantages of both logic programming and object-oriented programming paradigms. SLOT divides an application into many subtheories. Each theory has its own local database, and ways to interact with other theories. A local database is designed in such a way that it is described in Prolog and can be automatically translated into deductive and object-oriented databases. In our case, the target language is Chimera (Ceri and Manthey 1993) – a conceptual modeling language for the IDEA project.

We have used special SLOT theories for handling uncertainty and ambiguity data and theories for using inference rules. Users can interact with the system through a graphical use interface or by issuing Prolog form queries. Figure 5 provides a screen dump for a particular session. More details can be found in Jonker et al. (1994).

So far the proposed formalism has not yet been fully implemented in our database system. We have begun to investigate the implementation along two possible

lines. One involves using theorem proving techniques in which the reasoning process will take several steps. First, metric information is translated into topological relations. Then a special spatial logic theorem prover is employed to derive as much as possible new topological information in each local orientation frame. Since there are usually a few intervals (DNA fragments) in each local orientation frame, this process will be reasonably fast. After that, we merge local orientations, and then construct the global ordering.

We can also implement the proposed formalism along the line of qualitative simulation developed in Cui et al. (1992). One advantage of this approach is that distances can be directly used with the topological relations and no translation is required. The final results are usually a set of possible orders for the given intervals (DNA fragments in our case).

Conclusion

In this paper, we have discussed the problems in building a large deductive and object-oriented database to support molecular biology, especially problems related to represent and construct genome maps. We have described an integrated formalism based on a spatial and temporal logic, which extends Allen's interval logic to include metric information and local orientation. The binary relations form a relational hierarchy corresponding to the coarseness of our knowledge about the relationship between two DNA fragments. For example, unordered fragments are represented by the highest level relation; imprecisely ordered fragments by the middle relations and precisely ordered fragments by the leaf relations in the relational hierarchy.

Based on this integrated formalism, we have designed a deductive and object-oriented data model for molecular biology. We have partially implemented this in a prototype database system offered by the IDEA project, together with a formal object-oriented specification system SLOT (Jonker et al. 1994, Cui et al. 1993, Cui and Fox 1994) because the prototype database system does not yet support all the functionalities required. Our initial experience has shown the adequacy of our formalism for representing and reasoning with various source information in the context of Human Genome project.

There are clearly several areas that need to be investigated further. Our initial database needs to be expanded to include more data and this will suggest how we can improve our data model and inference mechanism. We will work on a composition table for combining local orientations in the same spirit as the transitivity tables in spatial and temporal logic (Allen 1983, Randell et al. 1992, Gooday et al. 1994). In addition, the spatial logic we are using also supports splitting or gapped intervals. This could be used to describe loci that involve alternative splicing (Bordereaux et al. 1990) and protein coding messages produced by splicing. We also intend to extend the formalism to handle

uncertainty and conflicting experiment data, possibly along the line of argumentation by Fox et al. (1992). We would also like to design a general declarative query language for our genetic database.

Acknowledgements

The financial support from IDEA project is gratefully acknowledged. We thank Drs. Catherine Hearne, Simon Parsons and John Fox for their comments, suggestions and inspirations.

References

- Allen, J. F. 1983. Maintaining Knowledge about Temporal Intervals. *ACM Transactions on Database Systems*, 15(3).
- Allen, J. F. 1984. Towards a General Theory of Action and Time. *Artificial Intelligence* 23(2).
- Bordereaux, D.; Fichelson, S.; Tambourin, P.; and Gisselbrecht, S. 1990. Alternative splicing of the Evi-1 zinc finger gene generates mRNAs which differ by the number of zinc finger motifs. *Oncogene*, 5.
- Ceri, S.; and Manthey, R. 1993. First Specification of Chimera. IDEA.DE.2P.004.02.
- Clarke, B. L. 1981. A Calculus of Individuals Based on Connection. *Notre Dame Journal of Formal LOGIC*, Vol. 2, No. 3.
- Clarke, B. L. 1985. Individuals and Points. *Notre Dame Journal of Formal LOGIC*, Vol. 26, No. 1.
- Cohn, A. G. 1987. A More Expressive Formulation of Many Sorted Logic. *Journal of Automatic Reasoning*, Vol. 3 No. 2.
- Cui, Z.; Cohn, A. G.; and Randell, D. A. 1992. Qualitative Simulation Based On A Logical Formalism of Space and Time. In proceedings of AAAI92.
- Cui, Z.; Cohn, A. G.; and Randell, D. A. 1993. Qualitative and Topological Relationships in Spatial Databases. *Advances in Spatial Databases*, eds. D. Abel and B. C. Ooi, Lecture Notes in Computer Science 692.
- Cui, Z.; and Fox, J. 1994. A General Framework for Building Intelligent Database Applications. *to appear in the proceedings of the 3rd Golden West International Conference on Intelligent Systems*, Las Vegas.
- Cui, Z.; Fox, J.; and Hearne, C. 1993. Knowledge based Systems for Molecular Biology: the Role of Advanced Database Technology and Formal Specification. In Proceedings of IJCAI93 workshop on AI and Genome, Chambery.
- Fox, J.; Krause, P.; and Ambler, S. 1992. Arguments, Contradictions and Practical Reasoning. in Proceedings of ECAI92, Vienna.
- Freksa, C. 1992. Temporal Reasoning Based on Semi-Intervals. *Artificial Intelligence*, 54, pp. 199-227.
- Gooday, J. M.; Bennett, B.; and Cohn, A. G. 1994. Efficient Generation of Composition Tables and Their

- Compaction. Technical Report, School of Computer Studies, Leeds University.
- Graves, M. 1993. Integrating Order and Distance Relationships from Heterogeneous Maps. In Proceedings of ISMB93.
- Guidi, J. N.; and Roderick, T. H. 1993. Inference of Order in Genetic Systems", In Proceedings of ISMB93.
- Han, K.; and Noordewier, M. 1992. An Efficient Algorithm for Ordering Random Genomic Clones. Tech. Report, Computer Science Department, Rutgers University.
- Hearne, C; Cui, Z.; Parsons, S; and Hajnal. 1994. Prototyping a Genetics Deductive Database", in Proceedings of ISMB94, this volume.
- Honda, S.; Parrott, N. W.; Smith, R.; and Lawrence, C. 1993. An Object Model for Genome Information at All Levels of Resolution. In Proceedings of ISMB93.
- Jonker, W.; Bressan, S.; Sikeler, A; Hearne, C; Parsons, S.; and Cui, Z. 1994. On the Use of IDEA Database Technology in Molecular Biology. Technical Report, Imperial Cancer Research Fund, London.
- Kautz, H. A.; and Ladkin, P. B. 1991. Integrating Metric and Qualitative Temporal Reasoning. In proceedings of AAAI91, Anaheim, CA.
- Kazic, T.; Lusk, E.; Olson, R.; Overbeek, R.; and Tuecke, S. 1990. Prototyping Databases in Prolog. The Practice of Prolog, Leon Sterling, ed. MIT Press.
- Letovsky, S.; and Berlyn, M. B. 1992. CPROP: A Rule-Based Program for Constructing Genetic Maps. *Genomics* 12, pp435-446.
- Mott, R.; Grigoriev, A.; Maier, E.; Hoheisel, J.; and Lehrach, H. 1993. Algorithm and Software Tools for Ordering Clone Libraries: application to the mapping of the genome of *Schizosaccharomyces pombe*. *Nuc. Acids Res.* 21, pp1965-1974.
- Olsen, M.; Hood, L.; Cantor, C.; and Botstein, D. 1989. A Common Language for Physical Mapping of the Human Genome. *Science* 245, pp. 1434 - 1435.
- Pecherer, R. M. 1993. Contig Graph Tool: A Graphical Interface for Contig Physical Map Assembly. In proceedings of Hawaii conf. of system sciences.
- Randell, D. A.; and Cohn, A. G. 1992. Exploiting Lattice in a Theory of Space and Time. *Computers and Mathematics with Applications*.
- Randell, D. A.; Cui, Z.; and Cohn, A. G. 1992. A Spatial Logic based on Regions and Connection. In Proceedings of the 3rd Int Conf on the Principles of Knowledge Representation and Reasoning, Morgan Kaufmann.
- Yoshida, K.; Smith, C.; Kazic, T.; Michaels, G.; Taylor, R.; Zawada, D.; Hagstrom, R.; and Overbeek, R. 1992. Toward a Human Genome Encyclopaedia. In Proceedings of Int. Conf. Fifth Gen. Comp. Systems, 307, ICOT.