standalone tools with a common data format under a single, standalone, or web-based
interface, to integrative, distributed, and extensible bioinformatics workflow management
systems. The field of bioinformatics, with its development of many kinds of freeware, has been
rather chaotic. One explanation for this problem is that bioinformatics is a very large field,
including many different knowledge domains, each with its own practitioners working in
isolation from each other. It is only recently that these diverse areas have been brought
together under the same umbrella of bioinformatics. Later, we will suggest that we are
currently in a stage of “sifting out” what is really needed and that future years will bring
a more or less standard data analysis methodology, involving data mining and text mining.
But for now, Table 15.5 provides a list of some of the software programs that are being
used in the field of bioinformatics. BLAST, which is the major database search tool, and
ClustalW2, which is the major multiple alignment of sequences tool, were discussed
previously so are not included in the table.

**TABLE 15.5 Important Software Programs Used in Bioinformatics***

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>URL</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>GenScan</td>
<td><a href="http://genes.mit.edu">http://genes.mit.edu</a></td>
<td>Prediction of genes, e.g., DNA</td>
</tr>
<tr>
<td>Prediction</td>
<td>PsiPred</td>
<td><a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a></td>
<td>Prediction of protein structure</td>
</tr>
<tr>
<td>Prediction</td>
<td>Mfold</td>
<td><a href="http://www.bioinfo.rpi.edu/applications/">www.bioinfo.rpi.edu/applications/</a></td>
<td>Prediction of RNA structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mfold/</td>
<td></td>
</tr>
<tr>
<td>Visualization</td>
<td>Logos</td>
<td><a href="http://weblogo.berkeley.edu">http://weblogo.berkeley.edu</a></td>
<td>MSA (=Multiple Sequence Alignment) visualization</td>
</tr>
<tr>
<td>Visualization</td>
<td>Rasmol</td>
<td><a href="http://www.umass.edu/microbio/rasmol/">www.umass.edu/microbio/rasmol/</a></td>
<td>Visualization of structure</td>
</tr>
</tbody>
</table>

*BLAST and ClustalW2 are also very important but were discussed earlier in this chapter.

HOW DO WE APPLY DATA MINING METHODS TO BIOINFORMATICS?

There is a multitude of open source and freeware computer software sources on the
Web, which provide programs written by various bioinformatics researchers. Addition-
ally, some have written books explaining how users can write their own programs. One
of these is with the use of Perl, available as an open source code base. Tisdall (2001) wrote
*Beginning Perl for Bioinformatics*, in which sequences examples are emphasized. Perl is a
relatively easy computer language to learn, and we mentioned it in Chapter 9 on text
mining. Tisdall (2001) shows how to build an interface for the NIH GenBank database,
which is mentioned earlier in this chapter. He also provides an introduction on how to
use Perl to interface with the BLAST sequence alignment tool, mentioned earlier in this
chapter. It allows the user to develop skills in using Perl to parse annotations in GenBank
and BLAST output.