Journal of Computer-Aided Molecular Design, **16:** 809–818, 2002.
KLUWER/ESCOM

Distant collaboration in drug discovery: The LINK3D project

Manuel Pastor[1]*, Paolo Benedetti[2], Angelo Carotti[3], Antonio Carrieri[3], Carlos Díaz[1], Cristina Herráiz[1], Hans-Dieter Höltje[4], M. Isabel Loza[5], Tudor Oprea[6], Fernando Padín[5], Francesc Pubill[7], Ferran Sanz[1], Friederike Stoll[4] & the LINK3D Consortium[8]

[1]*GRIB, IMIM/UPF, C/ Dr. Aiguader 80, E-08003, Barcelona (Spain);* [2]*MIA srl, Perugia (Italy);* [3]*Università degli Studi di Bari (Italy);* [4]*Heinrich-Heine-Universität Düsseldorf (Germany);* [5]*Univ. de Santiago de Compostela (Spain);* [6]*ASTRAZENECA AB , Mölndal (Sweden);* [7]*LACER SA, Barcelona (Spain);* [8]*In addition to the mentioned institutions the LINK3D Consortium is constituted by AL Digital Ltd, London (UK); Dep. of Technology, UPF, Barcelona (Spain) and the EC – Joint Research Centre, Ispra (Italy).*

Received 10 October 2002; Accepted 10 January 2003

Key words: drug discovery, e-work, Internet, knowledge sharing, synchronous telecollaboration.

Summary

The work describes the development of novel software supporting synchronous distant collaboration between scientists involved in drug discovery and development projects. The program allows to visualize and share data as well as to interact in real time using standard intranets and Internet resources. Direct visualization of 2D and 3D molecular structures is supported and original tools for facilitating remote discussion have been integrated. The software is multiplatform (MS-Windows, SGI-IRIX, Linux), allowing for a seamless integration of heterogeneous working environments. The project aims to support collaboration both within and between academic and industrial institutions. Since confidentiality is very important in some scenarios, special attention has been paid to security aspects. The article presents the research carried out to gather the requirements of collaborative software in the field of drug discovery and development and describes the features of the first fully functional prototype obtained. Real-world testing activities carried out on this prototype in order to guarantee its adequacy in diverse environments are also described and discussed.

Introduction

Drug discovery and development (DDD) projects usually involve hundreds of researchers, often working in centers scattered over different countries. Recent merging of large pharmaceutical corporations and the tendency to externalize relevant research tasks to contract research organizations (CROs) and academic organisations contribute also to distributed research. In this scenario, efficient communication and coordination between the researchers involved in DDD projects is of capital importance.

Nowadays, the advance of Information Technologies (IT) and the widespread use of the Internet have contributed to break geographical barriers. Scientists working in any knowledge field use the Internet in their everyday work to communicate and to access remote data. Often, closely collaborating researchers use data repositories, accessible through web pages or file transfer protocol (FTP) servers, and use e-mail to exchange messages. These resources can be classified as 'asynchronous' since the data is deposited and accessed at different times. A more sophisticated category of resources, the so-called 'synchronous', permits the remote, real-time interaction of two or more people. The idea of using the Internet for synchronous telecollaboration is not new and articles about this subject can be traced back to the first days of Internet [1]. Unfortunately, this early enthusiasm did not succeed to produce broadly used collaboration tools.

Recently, the potential of Internet to enhance collaboration in scientific research has been revisited and

*To whom correspondence should be addressed.
E-mail: mpastor@imim.es

reviewed [2], presenting some successful results in diverse fields. DDD is an area particularly well suited for the use of such collaboration tools, for a number of reasons: (i) the computer literacy level of the professionals involved is particularly high, (ii) the teams are often large and rather disperse, (iii) much research is carried out in corporate environments eager to invest in resources increasing the efficiency of the research, etc. As in other fields, the implementation of synchronous collaborative tools can have a positive impact in terms of increasing the efficiency of the research, speed up of the whole discovery process, and promote academic-industrial partnerships. Reduction in the number of travels can produce significant savings in travel costs and increase the quality of life of the professionals involved.

For all the aforementioned reasons, the LINK3D project [3] was started with the final aim of producing tools supporting the application of telecollaboration practices in the DDD field. The project is being undertaken by a consortium of ten European institutions, both of industrial and academic nature (see Table 1), which incorporates the required expertise on software engineering and on the different drug discovery specialties needed to carry out testing in realistic scenarios.

The specific objectives of the project were: (i) to evaluate the requirements of synchronous collaboration tools in the field of DDD, (ii) to develop state-of-the-art software and (iii) to check its adequacy to the requirements in real-world environments. These three points have been used here to guide the presentation and discussion of the preliminary results of the project.

Results and discussion

User needs investigation

No adequate software can be produced without a careful investigation on the specific needs of its potential users. In this case, the research was expected to shed light on the particular requirements of synchronous collaboration tools addressed to DDD professionals. The investigation focused on a survey to prospective users aiming to obtain information about:

1. Technological issues: hardware and software currently used by DDD professionals, including operating systems, communication tools and Internet usage.
2. Application field characteristics and requirements: description of data formats used in DDD projects meetings; specific requirements; objects manipulation required for DDD projects.
3. Current collaborative practices and potential for LINK3D: average length and frequency of DDD projects meetings; cost; usage of other telecollaboration tools; security issues.

The survey was sent to a number of carefully chosen DDD professionals. The inclusion criteria specified that the professionals should be relatively senior (minimum 5 years of professional experience in the field) and currently or recently involved in collaborative DDD projects. Participants were selected from different laboratories and, preferably, organisations (no more than 3 from the same laboratory), trying to balance the mixture of academic and industrial respondents as well as the mixture of 'computational' and 'experimental' drug discovery professionals. The survey was completed in May 2001, resulting in a total of 72 questionnaires completed by professionals from 6 European countries. The main findings are summarised in Table 2.

Technological issues

With respect to the hardware and software tools used in DDD, the survey reveals the diversity of the platforms used: 90% of users are currently using MS-Windows and almost 70% use a UNIX-based system (mainly Linux and SGI-IRIX). Only 20% of users use Macintosh OS, mainly in academic environments. Therefore, in order to allow seamless integration into current platforms, the LINK3D software should run at least under three operating systems: MS-Windows, Linux and SGI-IRIX.

With respect to the Internet accessibility, more than 90% of the surveyed users declare to use the Internet in their daily work for browsing the web and for e-mail purposes, and more than 64% use FTP facilities. This is not surprising and is in agreement with our previous knowledge of the high level of Internet usage in the field.

Application field characteristics and requirements

The analysis of the survey results provided useful information on a large number of aspects relative to the desirable characteristics of the LINK3D software. Remarkably, high importance was given to the presence of a specialised whiteboard able to represent shared data. The representations and file formats perceived as more important were 2D molecular structures (sdf,

Table 1. Institutions participating in the LINK3D Consortium.

Participant name	Country	Institution type
Fundació IMIM	Spain	Academic
LACER, S.A.	Spain	Corporate
Heinrich-Heine Universität Düsseldorf	Germany	Academic
Università degli Studi di Bari	Italy	Academic
Universidade de Santiago de Compostela	Spain	Academic
ASTRAZENECA R&D MÖLNDAL	Sweden	Corporate
M.I.A. Multivariate Infometric Analysis Srl	Italy	Corporate
Joint Research Centre – European Commission	Italy	Non-profit
A.L. Digital Ltd.	UK	Corporate
Universitat Pompeu Fabra	Spain	Academic

Table 2. Main findings of the user requirements investigation.

Technological issues	High diversity in the platforms (MS-Windows, Linux, SGI-IRIX). The use of Internet is very common (WEB browsing, e-mail and FTP).
Application field characteristics and requirements	The most important formats are 2D > 3D > generic text and graphics. – 2D: sdf, SMILES – 3D: pdb, mol2 – Other: html, jpg The most essential feature is a shared whiteboard. The most essential collaborative tool is a shared pointer, followed by marking, writing and rotating/moving/zooming.
Current collaborative practices and potential for LINK3D	DD projects last an average of 24 months with meetings every 2–6 months, with an average cost of 1000 Euro/attendant per meeting. Meetings often imply more than 2 attendants. 70% of users have never used teleconferencing or synchronous collaborating tools but express high interest in using LINK3D, expecting to replace about half of their meetings. High level of concern about confidentiality in the transmission.

SMILES) followed by 3D molecular structures (pdb and mol2) and generic text and graphics (html and jpg). Proprietary formats like MS-Word (doc), MS-PowerPoint (ppt) and Adobe Acrobat (pdf) were also given high importance, probably because these are commonly used in physical meetings (i.e. ppt) or together with asynchronous collaboration tools (i.e. doc and pdf as e-mail attachments).

With respect to the interaction tools that should be incorporated into the whiteboard, the ability to point to objects is regarded as the most essential, although the possibility to mark/write on objects and rotate/move/zoom them are also considered important functions. The ability to modify shared graphical objects is perceived by prospective users as the least essential.

The survey indicates that the software should incorporate highly specialised visualization and manipulation tools supporting a large number of formats. Object manipulation tools must be included, allowing the user at least to point to and to mark/write into the graphics.

Current collaborative practices and potential for LINK3D

The typical DDD project, as reported by users, is approximately 24-months long and implies non-local meetings on a bimonthly to half-yearly basis, each meeting costing around 1000 Euro per attendant. Often, more than two people participate in such meetings, what indicates that implementation of multicasting communications is important. Survey respondents were also concerned by aspects related with session

moderation and floor access control during a meeting and many of them requested implementation of moderation mechanisms.

An interesting result is the low level of experience among users when it comes to teleconference or telecollaboration tools: almost 70% of the users have never used such facilities. This finding confirms the initial assumption of a low level of use of synchronous technologies in the field and the opportunities of adequate software in that respect. Indeed, surveyed users are fairly interested in using the LINK3D system to replace some of their current face-to-face meetings. In average, users estimated that 45.5% of their current meetings could be held using software like the one envisaged in the framework of LINK3D.

All users declared to be very concerned with confidentiality in all the electronic communications related to their drug discovery projects. Computer-based professionals seem to be more concerned than experimental users, probably as a consequence of a greater awareness on the existing range of potential threats posed to the electronic communication of confidential information. From this answer it is clear that the software must guarantee the confidentiality of the discussion and the security of the data transmitted.

This survey was complemented with detailed passive observation of current practices (video recording or transcription of real DDD meetings) and interviews with experts of different academic and corporate institutions of the consortium (Table 1).

Not all specifications were determined by the results of the survey and some specifications of the products were determined at the very beginning of the project. For example, when developing synchronous collaboration tools, two alternative approaches were possible: to develop software taking advantage of special resources (high bandwidth or multicast networks, special hardware, etc.) or to develop software for already existing infrastructures. From the beginning, the latter was chosen, in order to make the program accessible to a wider range of potential users. One of the essential design specifications was therefore to build an application that was not too demanding with respect to bandwidth. Unfortunately, this decision conflicted with the demand of many prospective users to have video communication within the application. At the end video was discarded as a result of the observation of real DDD meetings, which showed how discussions are highly focused on molecular graphics and how a proper representation and manipulation of the whiteboard is by far more important than the gestural information provided by video.

System development

The project aims to obtain state of the art software in a novel field and therefore it was strictly required that a professional software engineering method was applied. The model chosen for development was the spiral model [4]: the project started gathering the user requirements and follows with iterative cycles of development and testing until a product with the appropriate level of quality is obtained. This method has proved very effective for developing complex applications in new areas. Moreover, the constant feedback obtained from users allows for refining the user interaction paradigms and often leads to more user-friendly applications.

At the moment of writing this manuscript, the third iteration of development is being completed and a full-featured prototype is available. Figure 1 shows a snapshot of the application interface.

One of the major problems was the need to run the application on several platforms. The obvious alternative to platform-dependent programming was to use a portable programming language like JAVA or C++ together with a multiplatform Graphic User Interface (GUI) toolkit like Qt [5]. This option was selected because it provided better integration with OpenInventor [6], the programming toolkit used for programming 3D interactive graphics as well as other components.

Communications follow the simple client-server organization shown in Figure 2. During a session, there is an ‘active’ token that passes from user to user following the floor-control rules. Only the active user is allowed to send data to a light server, which in turns broadcasts the data to the passive users. This method provides a cheap multicasting solution and integrates well with floor-control mechanisms implemented in the software (manual and automatic moderation).

Files used in the discussion by any partner are actually transmitted and stored in the virtual memory of the local workstations. The component in charge of managing the files (the so-called ‘repository’) lists them in the interface, in the form of a tree from where the user can pick any of them individually. The selected file can be sent to the shared whiteboard (‘shareboard’) for being displayed or saved to the local disks of other users if the owner of the file allows doing so. Files can be loaded into the repository before or during the discussion. In order to simplify this operation, drag&drop

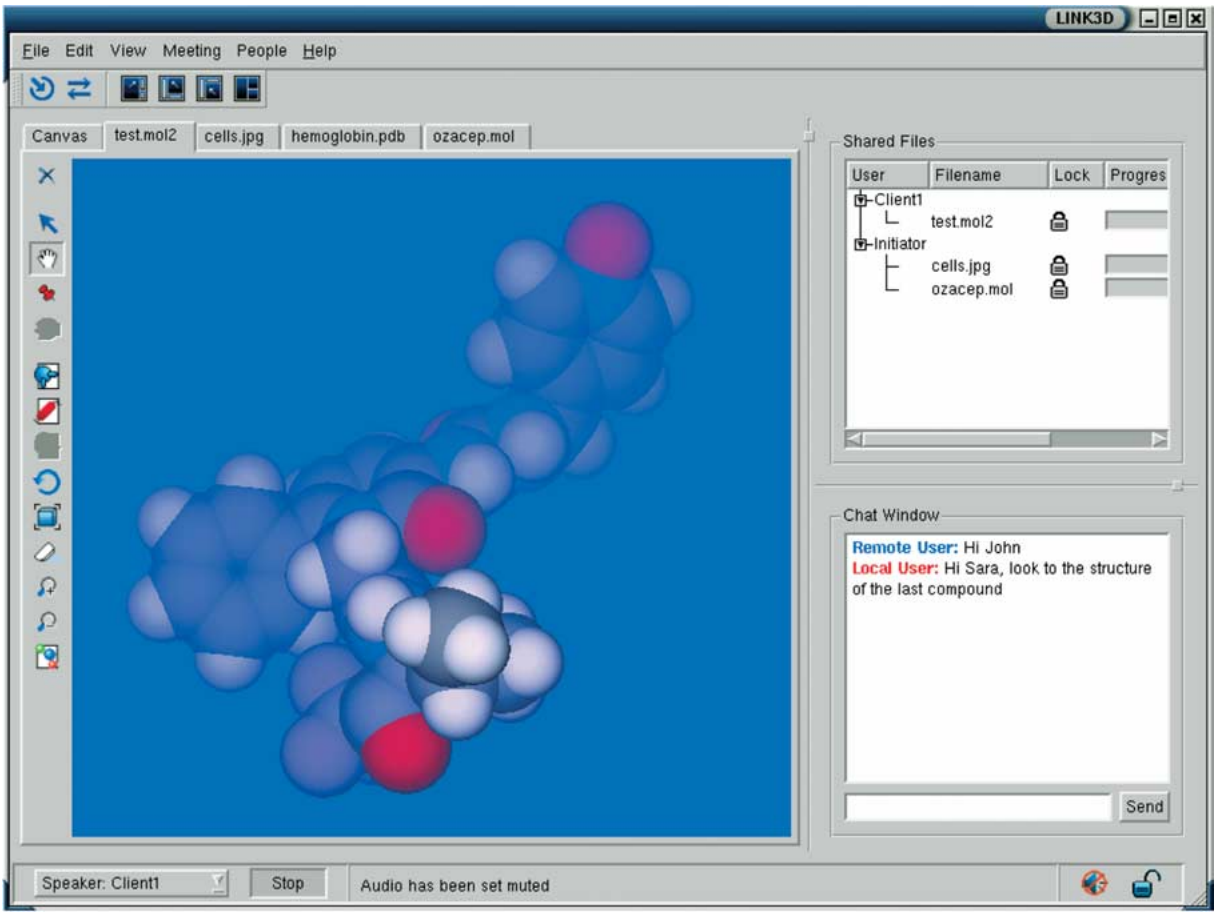

Figure 1. General aspect of the LINK3D graphic interface.

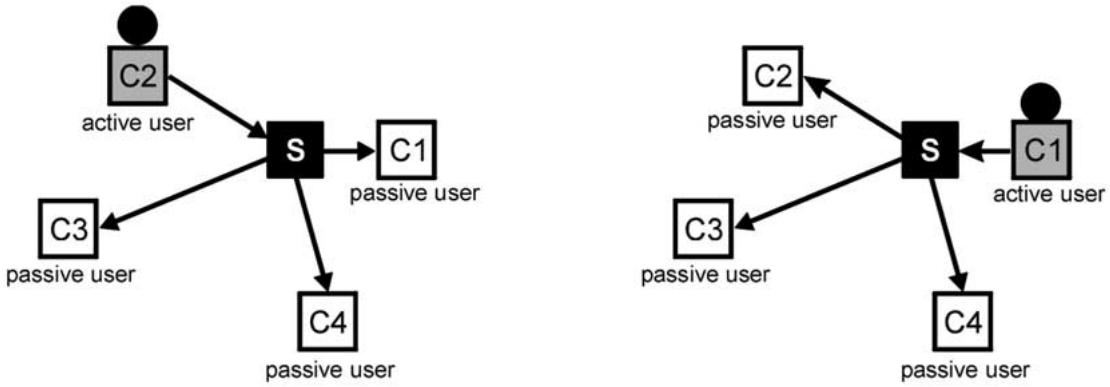

Figure 2. Diagram of data broadcast mechanism used in a LINK3D virtual meeting. The black circle represents the 'active' token that can pass from one user to another following floor-control rules. On the figure on the left hand side, the user C2 owns the token. In the figure on the right hand side, the token has been given to C1, which then becomes the active user. Only the client owning the token can send data to the server (represented by the filled square with an 'S'), which simply broadcasts the data to the passive clients.

interaction mechanisms were implemented and proved very practical in our tests.

The application provides also short text messages exchange (chat) plus full duplex audio. The audio is compressed using advanced codecs (like GSM) in order to keep the bandwidth requirements of the application as low as possible.

As can be seen in Figure 1, most of the interface of the application is used by the shareboard. During a meeting, the possibility of visualizing many files simultaneously has to be provided. Accordingly, the shareboard has been organized as a window with tabs, which can accommodate many layers piled one on top of each other and where users can change the layer being shown by simply clicking on the corresponding tab. There are different layer types. In the current version of the LINK3D software, there are four types of layers (Table 3), corresponding to the different kinds of data frequently used in DDD.

Each layer includes a bar with several buttons, which give access to specific tools for the layer. A significant effort has been paid to maintain a compromise between the consistency of the interaction paradigms across different layers and their adaptation to the data peculiarities of each layer. It was found that if the tools provided in different layers were too different (layer-specific), the user would feel confused and unable to interact intuitively, but if the consistency of the tools was forced too much, probably they would not be adequate for supporting a real discussion. For example, in the current version, all layers share similar (but not identical) tools for supporting the remote discussion.

During virtual meetings, the shareboard follows what we call the WYSIWES (What You See Is What Everybody Sees) paradigm. Any interaction, movement, style change, etc. is transmitted to make sure that all participants share exactly the same graphical view for supporting the discussions. This is an essential characteristic of the software, because no remote discussion can take place if the users are unsure about the consistency of their views. This is not easy to guarantee due to the limitations of the hardware (i.e. different monitor resolutions) and is one of the aspects still under active development.

layer3d

This layer provides a virtual 3D scenario where the user can load 3D molecular structures, isovolumes and, in general, any kind of 3D objects. The file formats supported are listed in Table 3. A single layer3d can hold any number of molecules, MIFs, etc., and they can be manipulated as a whole or individually. Moreover, the layer provides flexible mechanisms for selecting atoms, residues or regions of the molecules, either directly on the 3D object or from a hierarchical list, and editing their rendering style and color. For biomolecules, it is also possible to add simplified representations of their secondary structures and apply per-residue labeling. Figure 3 shows an example snapshot of the layer3d showing some of these features.

The scenario is interactive and users can manipulate 3D objects using the paradigm of a moving camera pointing to a static scenario. At any moment, the user having the control can rotate, translate and scale the molecules by dragging the mouse with one or several of its buttons pressed, in a similar way to what is done in the most popular molecular modeling packages. Such changes are reflected in the local screen and synchronously transmitted to the screen of all the computers connected. The position of the cursor in the screen of the user having the control is also transmitted and indicated in all the remote computers. As stated above, the idea is that all the participants share exactly the same graphical representation of the objects used in the discussion.

When discussing about 3D objects, it is very important to have tools permitting to highlight a certain region. The layer3d transmits in real time the position of the cursor of the user having the active token and this allows this user to interactively point to parts of the objects during the discussion. If a more permanent highlighting is required, it is possible to add ‘marks’. In the layer 3D, marks are simple geometrical shapes (spheres, cubes and cones), which can be inserted at the position of any atom present in the molecule and then moved in any direction using standard OpenInventor manipulators [6]. This mechanism offers a flexible and simple method for highlighting one or several molecular regions to be univocally referenced during the discussion, using common language expressions like ‘the region near the green cube’ or ‘the atom overlapped by the yellow cone’. This is an important feature of the present software since, when dealing with chemical entities, the inability of some of the participants to use proper chemical nomenclature can prevent them from participating, simply because they do not know how to tell other participants which part of the molecule they are referring to.

Table 3. Types of shareboard layers.

Name	Formats supported	Files per layer
layer3d	3D molecular structures (pdb, mol2, car, cor, kout) 3D models (wrl, iv) Molecular Fields (grd, cnt, kont, dat)	multiple
layer2d	2D molecular structures (mol, sdf)	multiple
layerbmp	Bitmap graphics (jpeg, png, bmp, tiff)	single
layertxt	Plain ASCII text (txt) HTML formatted text (html)	single

layer2d

The layer2d can load any number of 2D structures in any of the supported formats (see Table 3) and inserts them into a table, in which each row represents a single compound. 2D structures are represented in the first column; when the files contain extra information (i.e. biological activities, name, codes, molecular weight, etc.), this information is automatically added in the form of additional columns. This additional information can be used to reorganize the rows, simply by clicking on the column headers.

Supporting remote discussion about a table containing many compounds needs two levels of marking: marking a whole compound and marking a certain position of a compound. Accordingly, the layer2d provides two marking tools, one for whole rows and another for specific atoms. The second follows a mechanism similar to the one implemented in the layer3d and permits to insert colored geometrical figures (but 2D in this case) into the graphics. Figure 4 shows an example snapshot of the layer2d, showing a table with some structures and some atoms highlighted using both kinds of marks.

layerbmp and layertxt

The LINK3D software also provides a layer for representing bitmaps in the most extended formats (see Table 3), as well as a layer for displaying plain or HTML formatted text. Both layers support formats that are not specific of DDD discussions, but the experience has shown that they are very important as a backup method for sharing information not supported in a specific layer. For example, if during a discussion it is needed to show a pharmacology plot, it is often practical to take a snapshot of the screen and then load it into the bitmap layer. The layerbmp looks very much like the whiteboard that can be found in general purpose collaboration tools and allows to insert free-hand drawing and add text and marks. The layertxt supports plain text and simple HTML formatted documents and, in this case, the marker tools have been oriented to highlight lines and paragraphs.

At this point it would be worthy to compare the current LINK3D prototype with other similar software. Unfortunately, as stated above, collaborative software is not commonly used in DDD and there are only a few applications specifically developed for supporting remote discussion in this field. One of these applications is MICE (Molecular Interactive Collaborative Environment) [7], developed and maintained by the SDSC (San Diego Supercomputing Center). Basically, this application is oriented to structural biology applications and allows sharing 3D interactive models of proteins and other molecules. The application makes use of Java and Java3D technologies and can be used in most computer platforms. However, it is not specifically oriented to DDD and its applicability to this field is limited by its lack of support for other kind of molecular representations (i.e. 2D formulae) and other kinds of relevant data. MICE virtual meetings can only be held using dedicated servers at SDSC or similar centers and are rather oriented to academic users. In particular, the security of the communications is specifically not guaranteed. An alternative approach for setting up a collaborative environment is the use of generic collaborative tools, like MS-Netmeeting (for MS-Windows computers) or SGI-Inperson (for SGI workstations). These have the advantage of being inexpensive and of easy access for most users. Generic tools provide a very limited support for visualizing specialized data types but support high quality audio, video and offer the possibility to share any application running in one of the participating computers. This feature is very appealing because it allows collaboration using the same applications that the users apply in their everyday work. Unfortunately

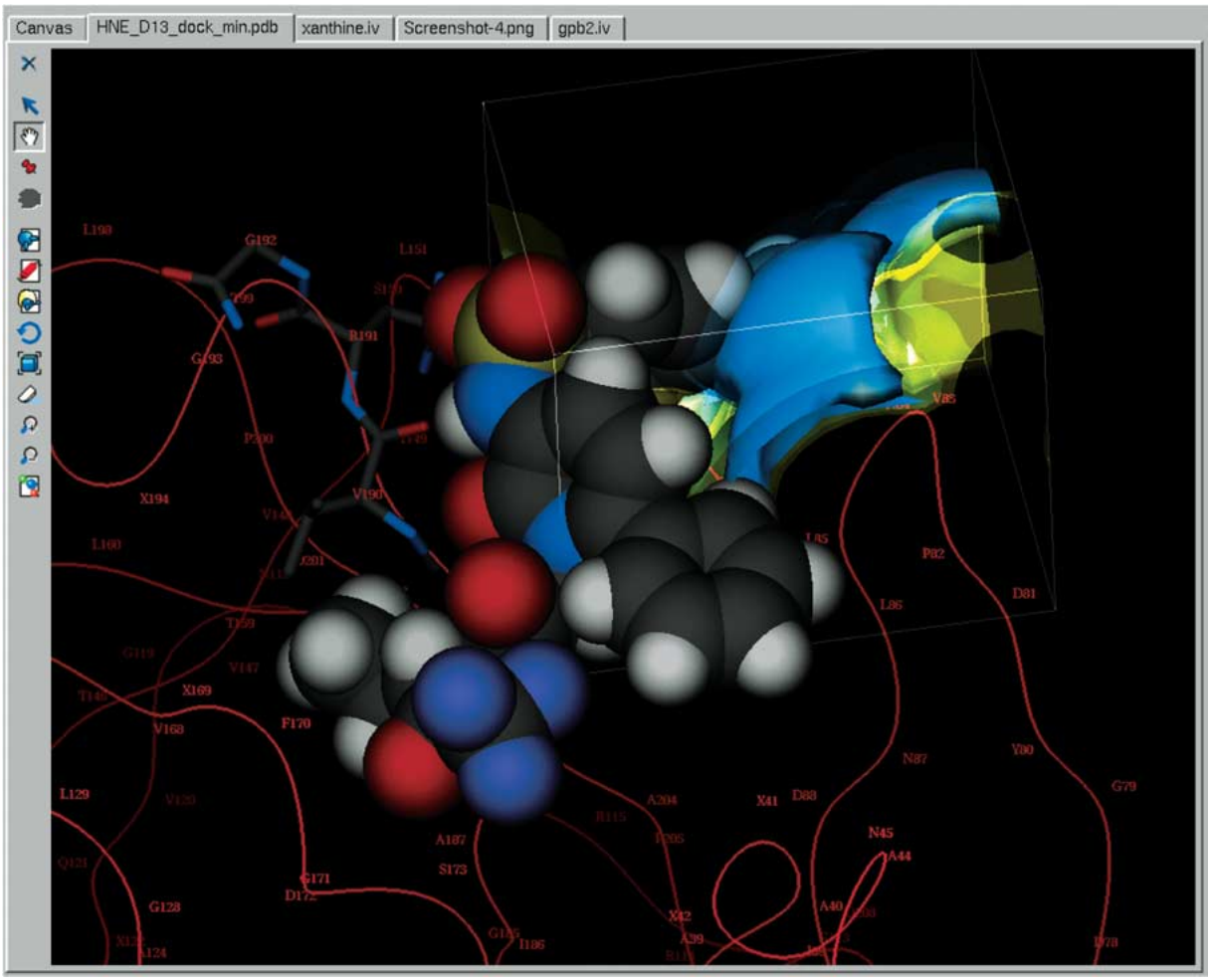

Figure 3. Snapshot of the layer3d. The scenario shows a protein represented only by a backbone line and the labels of the residues, a ligand rendered in CPK mode and a couple of isocontours representing different energy of interaction between the receptor and a chemical probe in the surroundings of the binding site.

this method is not practical for interactive 3D graphics, like those needed in molecular modeling, since the applications are shared by transmitting a continuous stream of compressed 'screenshots'. Therefore, when the molecules are moved, the remote representations exhibit sudden, abrupt movements that can be very disturbing for the discussion. Moreover, most generic tools have been developed for a single platform and their security capabilities are limited.

Real world testing

The iterative development model has allowed to test the LINK3D prototypes from early stages of the project and detect major design errors. Early testing results show a good acceptance by prospective users, and the application's performance seems to be satisfactory. According to the latest results, both the overall graphic quality and the speed achieved by the application were rated with an average 8 on a 0–10 scale. Special attention has been devoted to the interface. Users' feedback has been comprehensively sought in relation with this issue, and in particular regarding visible options and information on the screen, percentage of screen to devote to each specific section, interface structure, icons, menu bars, understanding of messages, dialog boxes, mouse buttons functionality, etc. Complementarily, video recordings of controlled groups of users interacting with the pre-prototypes were also analyzed and used for this purpose.

However, proper testing of this kind of software is not simple, because validation testing has special

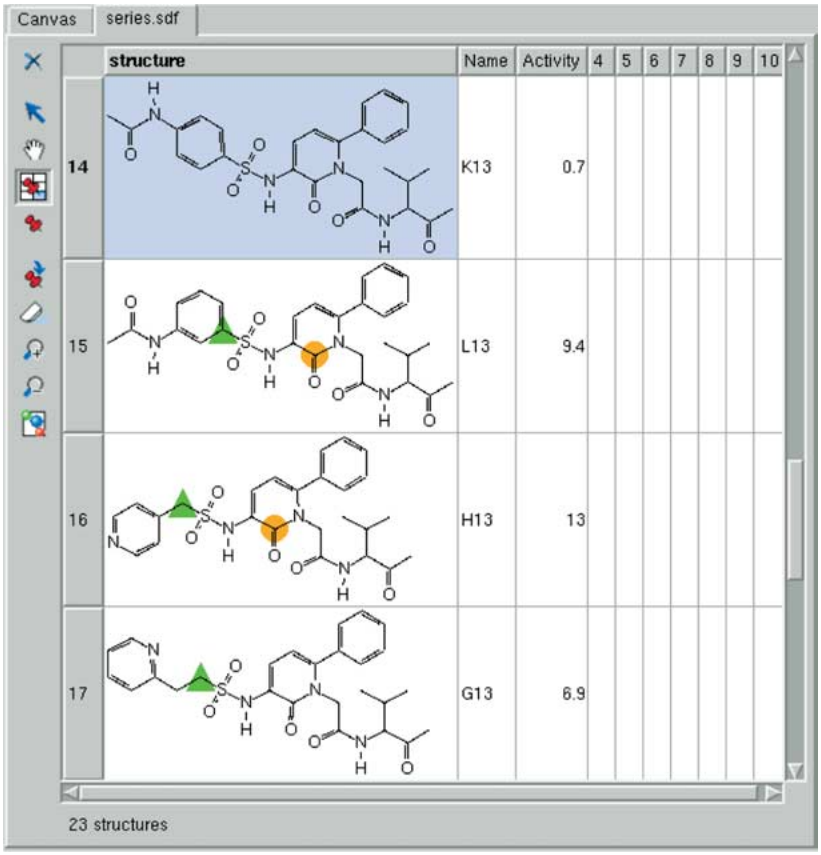

Figure 4. Snapshot of the layer2d. The table was built automatically from a single sdf file. The compound on top of the list has been marked. Also, some selected atoms have been assigned colored marks.

value when it is carried out in real-world environments and not only in simulated conditions. In this particular project, the potential application scenarios are diverse: it can be used to connect two separate sites of the same pharmaceutical company, two or more academic institutions collaborating in a joint project, an academic institution working as an external consultant for a pharmaceutical company, etc. Every scenario has its own peculiarities in terms of platforms more commonly used, peculiarities of the networks involved and security requirements.

A particular complex aspect of the project is to deal with the devices and policies set by institutions to restrict the access to and from the Internet. Indeed, the institutions interested in using LINK3D can follow very diverse protection policies: academic institutions tend to permit a more or less free access while private corporations protect incoming access with firewalls and outgoing communications with proxy servers. The only method to make sure that the connection strategies are practical is to perform extensive tests in real-world environments.

Fortunately, the LINK3D consortium includes partners representing the diverse user profiles (academic, corporate with many sites, small companies, etc.), which are participating in formally planned testing activities, involving setting up virtual meetings covering every potential utilization scenario.

Conclusions

The LINK3D software, being developed by a pan-European consortium, could be a highly valuable tool in DDD. It was designed after a careful investigation of the specific needs of DDD professionals. Some of its features (multiple platform support, extensive range of file formats supported, low bandwidth demand, specific collaboration tools, etc.) make them a unique product, highly differentiated from the general-purpose collaborative software that can be currently

found in the market. The product in itself is an example of successful collaboration between a team of professionals with highly diverse expertise (software development, security, drug design, drug synthesis and testing, etc.).

LINK3D aims to solve some of the bottlenecks currently affecting the efficiency of R+D processes that are crucial for competitiveness in the pharmaceutical area. When successfully completed, LINK3D will allow for a more efficient development of high added-value molecular entities, helping organisations in the field to exploit their expertise to the maximum advantage. The flexibility of the software will extend its potential use to other applications, like peer-to-peer meetings, remote external consultancy, aid to decision-making, remote educational guidance, etc. Moreover, the methodology used for LINK3D can be applied to other fields requiring specific customized solutions for synchronous telecollaboration.

The LINK3D software will be made available to non-profit institutions free of charge upon request. More information will be provided on the project home page at <http://www.tecn.upf.es/prj/link3d/>.

Acknowledgements

The LINK3D project is partially funded by the European Commission (Contract IST-2000-26338). We are also grateful to all the professionals who voluntarily answered our surveys, providing us with a source of highly valuable information.

References

1. Frivold, T.J., Lang, R.E. and Fong, M.W., The Second International WWW Conference'94: Mosaic and the Web. Chicago, IL, U.S.A., October 1994.
2. Teasley, S. and Wolinsky, S., Science, 292 (2001) 2254.
3. LINK3D: Linking organizations through a secure environment for distributed drug discovery. EU-funded project IST-2000-26338.
4. Cockburn, A. Agile Software Development. Addison-Wesley, Boston, 2002.
5. Qt 3.0, Trolltech AS (2001) http://www.trolltech.com.
6. OpenInventor 2.1, Silicon Graphics Inc. (2001) http://www.sgi.com.
7. MICE, San Diego Supercomputer Center, http://mice.sdsc.edu.