

An Evolutionary General Purpose WebGIS to Disclose EGFR Mutations in Lung Cancer

Domenico Aliotta, Pietro Buffa, and Gennaro Iaccarino

10th International Conference on Visual Information Systems (VISUAL08) -
Salerno, Italy

Molecular data visualization is considered a key aspect in the study of main biological processes and allow a fast acquisition of structural and functional information on different proteins of biomedical interest (such as EGFR).

Unfortunately, many data visualization tools offer a specific graphic vision for every kind of information that often cannot be easily integrated and possible correlations among them are not easily pinpointed.

The heterogeneity of these tools forces biologists to use them in different steps of their work, without correlations with results, attempt to obtain the desired information.



These difficulties necessitate the realization of a new tool (Web based), that combines all these features with a simple data management and shows results for a better scientific knowledge.

Geographic Information Systems (GISs) were born as systems for capturing, storing, analyzing, managing data and associated attributes which are spatially referenced to Earth.

Features are categorized separately and stored in different map layers, which share a common coordinate space.

In particular, layers can be added for locating measured data using an evolutionary concept.

This way of organizing data in a GIS, makes maps much more flexible to use since these layers can be combined in any useful manner.

Similarly, biological structures (proteins, genes, etc.) define a biological space coordinates, with the same characteristics of any other, thus perfectly performable with a GIS.

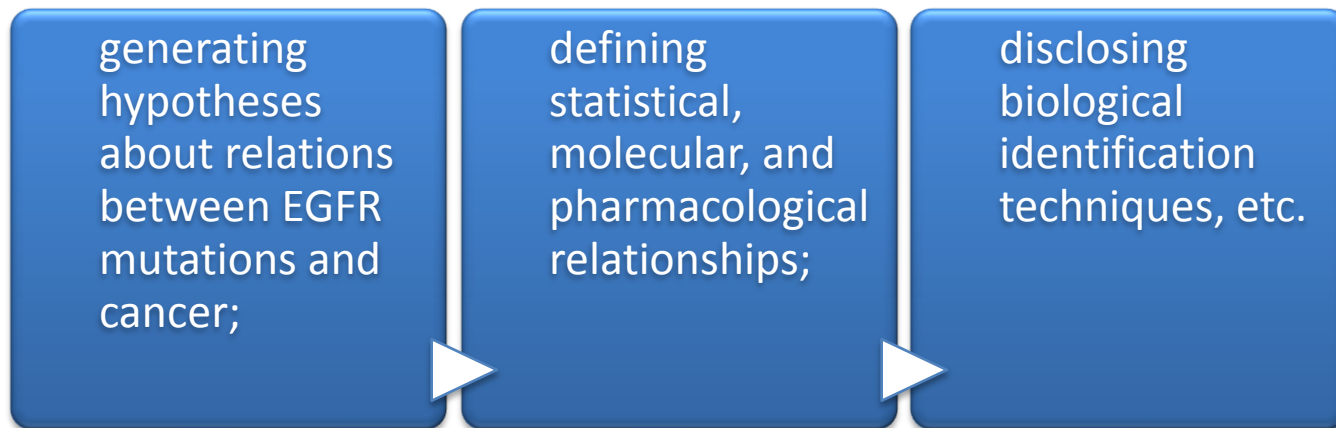
One of the first example of spatial information systems, in which geographic informations represent genome data, was developer by Dolan et al. This system was named GenoSIS (Genome Spatial Information System), and it was an application that used the concepts and tools of geographic informations science for the interpretation and modeling of genome data but it was not Web Based. It represented a novel approach to several problems of current interest to molecular geneticists.

As case of studie for our system we show relationship between the most common mutations in EGFR (Epidermal Growth Factor Receptor) kinase domain and lung cancer, that accounts for one third of all deaths from cancer worldwide

It has been widely demonstrated that EGFR is implicated in non-small-cell lung cancer (NSCLC), comprising 80% of lung cancers, glioblastoma, and breast cancer, where its oncogenic potential is stimulated by protein overexpression or by somatic “gain-of-function” mutations.

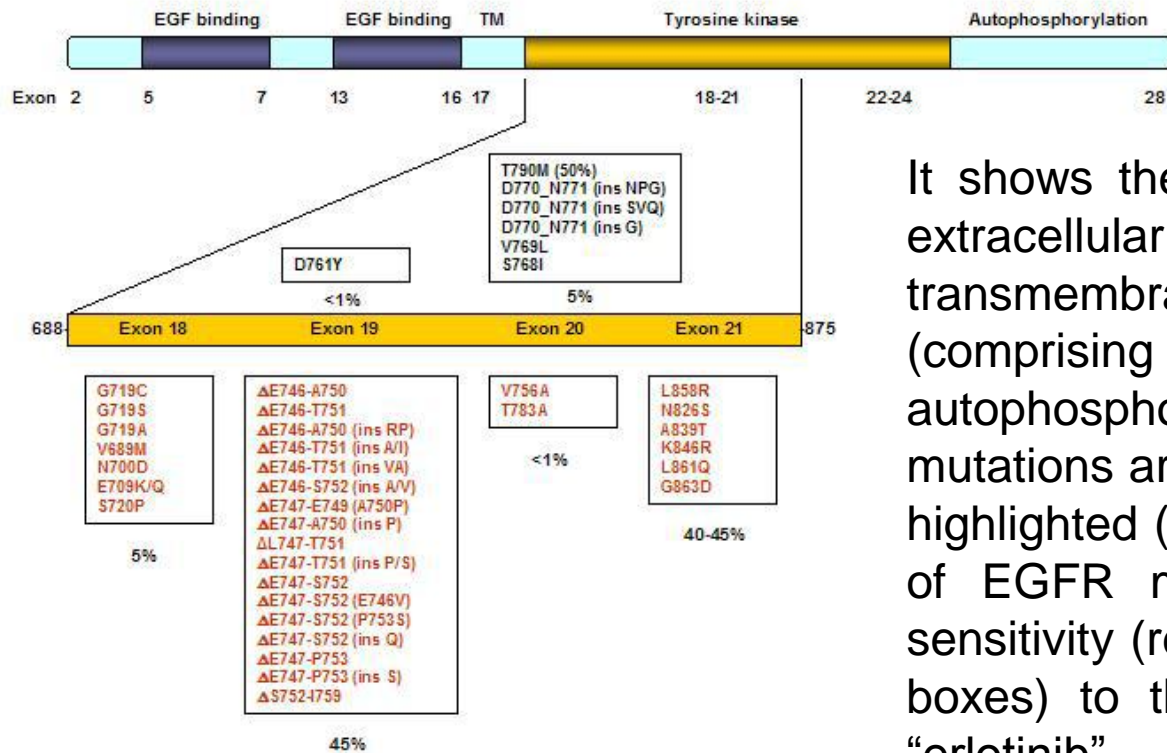
We present an evolutionary general purpose G.I.S. in which geographic information is replaced by biological data representing the main mutation regions within the kinase domain of EGFR.

Our system can be used for:



So we show that GIS capabilities can be used not only for geographic items, but for any kind of data, as a general purpose system.

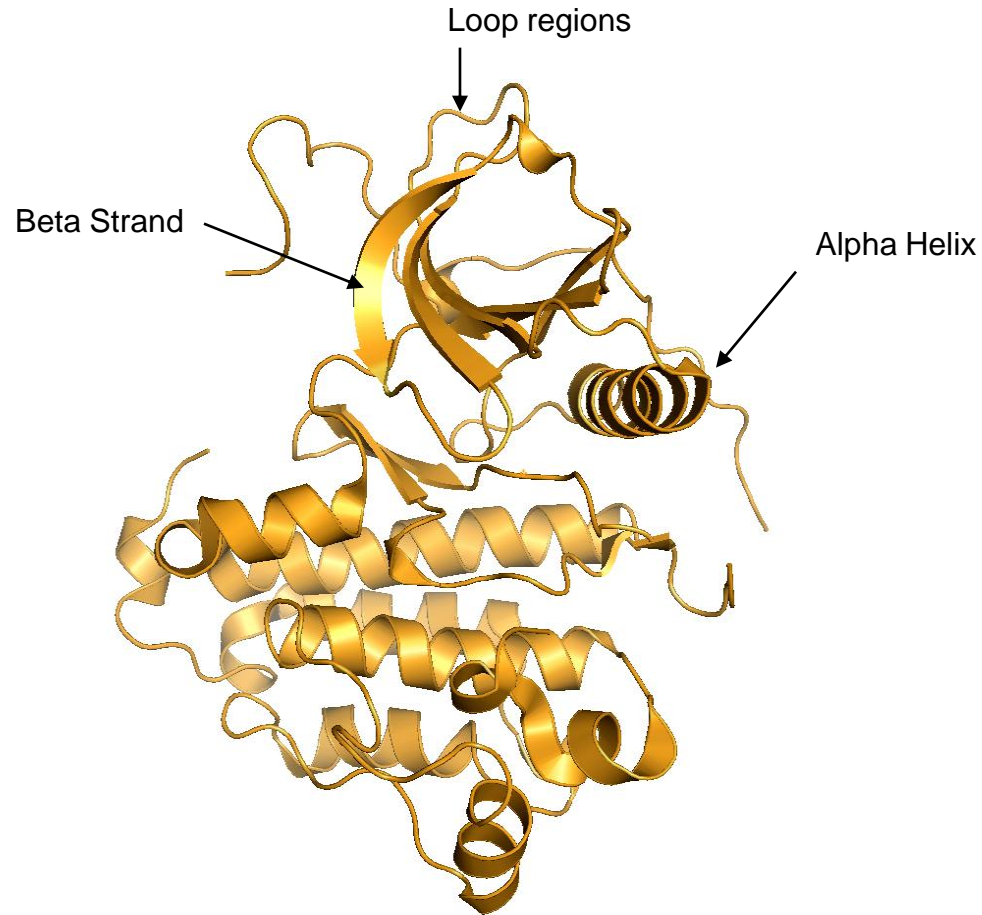
Representation of the Epidermal Growth Factor Receptor (EGFR).



It shows the distribution of exons in the extracellular (EGF binding), transmembrane and intracellular domains (comprising the tyrosine kinase and autophosphorylation regions). Relevant mutations are situated in Exons 18/21, are highlighted (yellow bar); and a detailed list of EGFR mutations is associated with sensitivity (red boxes) or resistance (black boxes) to the main drugs, “gefitinib” or “erlotinib”. Moreover, frequencies of mutations are shown.

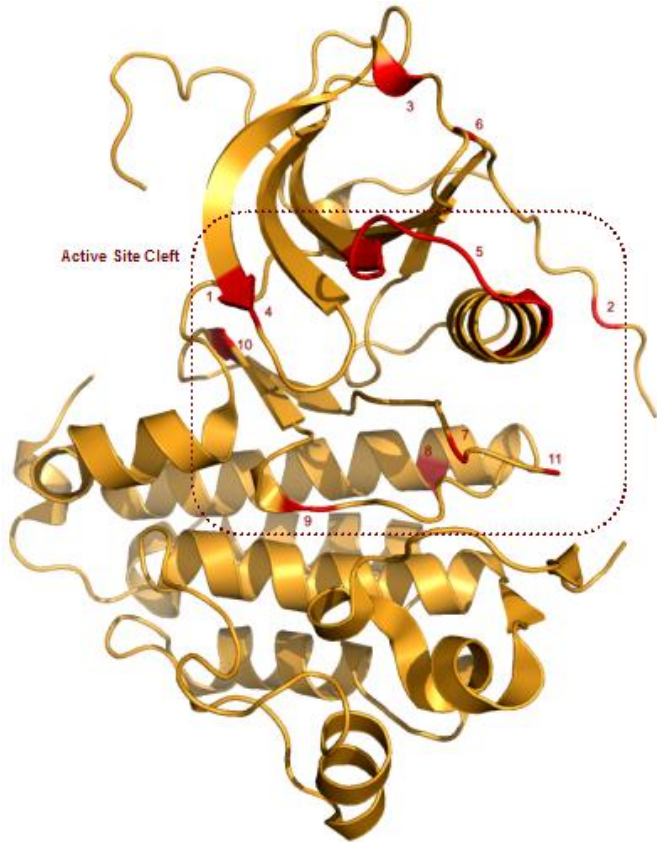
Epidermal Growth Factor Receptor and Lung Cancer

EGFR kinase domain has been intensely studied in its 3-Dimensional domain organization.



Epidermal Growth Factor Receptor and Lung Cancer

A number of distinct mutations have been structurally identified around the active site cleft of the EGFR kinase domain (EGFR-KD). The protein regions more frequently subject to mutations are shown in red.



Our WebGIS prototype uses these mutations as example data set.

We plan on dividing a protein in four fundamental levels of increasing complexity:

- 1) Amino acid sequence level;
- 2) Secondary structure organization level;
- 3) Domains organization level;
- 4) 3-Dimensional organization level.

All related information of each level are show into the GIS System.

System Overview

We used GIS's strengths in order to improve 2D biological models and link them using multilayer, so it is possible to produce biological maps, that can be enriched with new layers, data, simulations, statistics and so on.

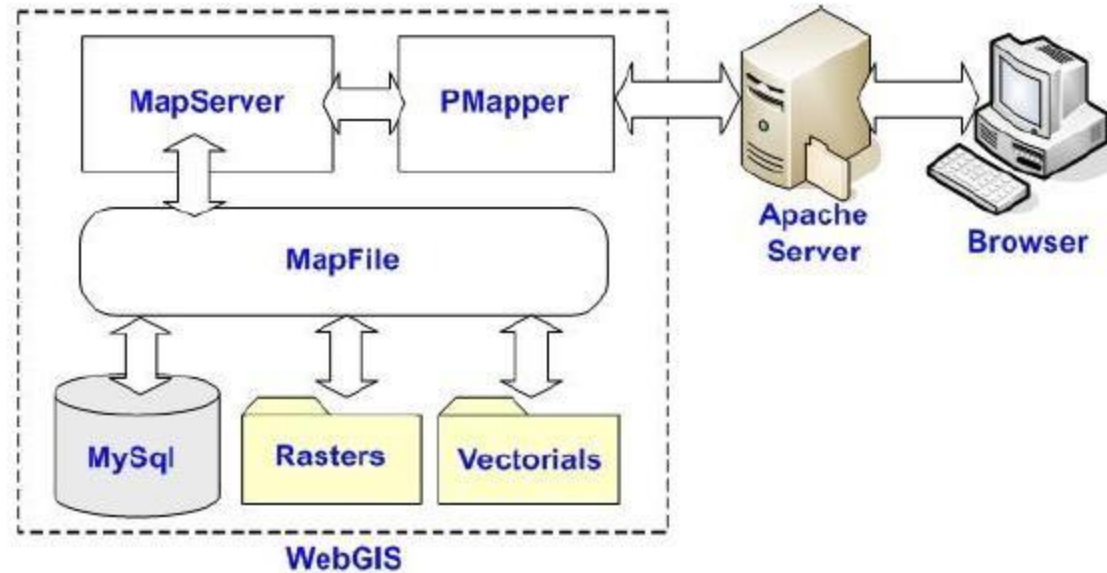
Our goal was to show that general purpose GIS extends the concept of Geographical Information System, increasing usability for each kind of data, also providing a central scientific archive on the web.

Our WebGIS infrastructure is based on an original modification and integration of several major open source software systems for

Data management (MySQL),
Geographical data visualization through internet (MapServer),
Geographical data analysis (Quantum GIS),
and Internet software management (Apache).



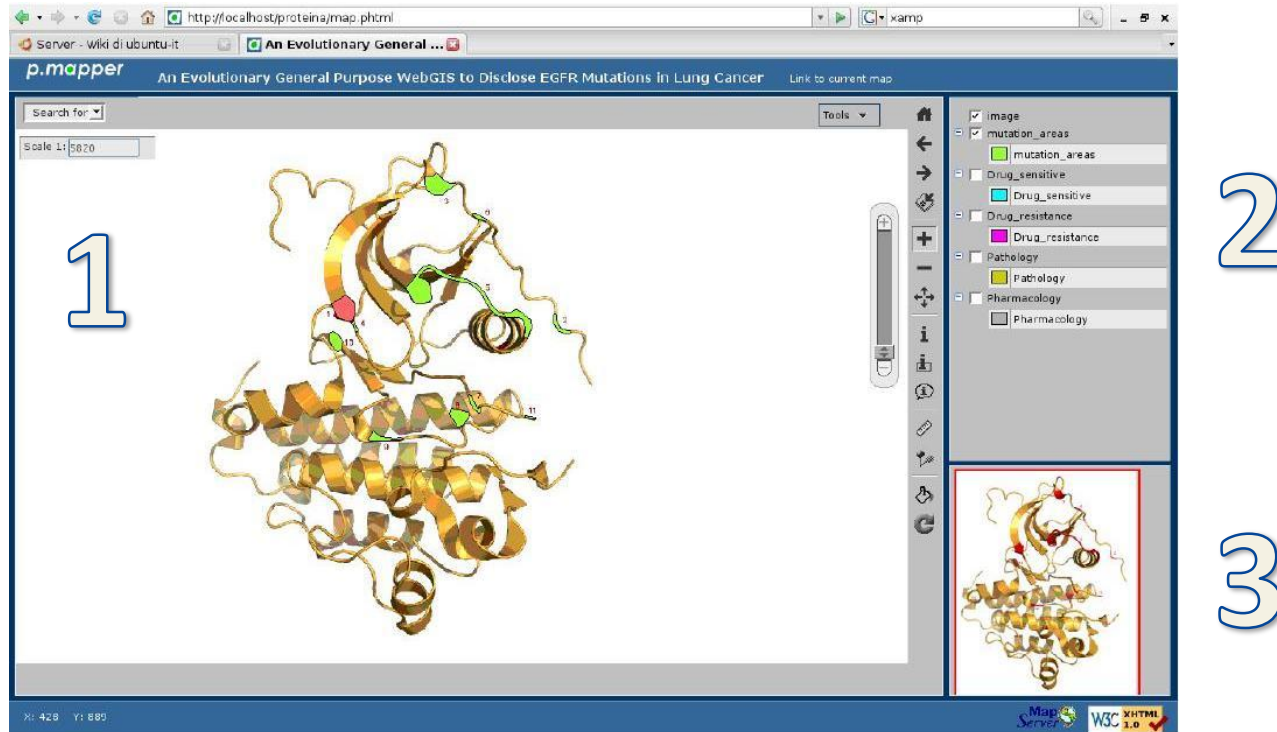
System Overview



The system can be accessed by simply using an Internet browser or a desktop GIS client. The Apache Web Server receives and processes HTTP requests, and invokes MapServer (through PMapper infrastructure). The MapFile includes information about data (rasters and vectorials) and is the system core.

It is processed by MapServer and sent to the HTTP server (Apache) that will send it to the browser (client).

Main Window



GIS main window is divided into three basic areas:

- 1) a main map generated by MapServer, surrounded by various control instruments, that reproduces the 3D ribbon representation of EGFR with all the activated layers;
- 2) an interactive legend through which users can activate and deactivate layers;
- 3) An interactive reference map, which offers also the navigation of the protein.

Functionalities

The application offers several tools that can be activated using buttons in the pipe bar.

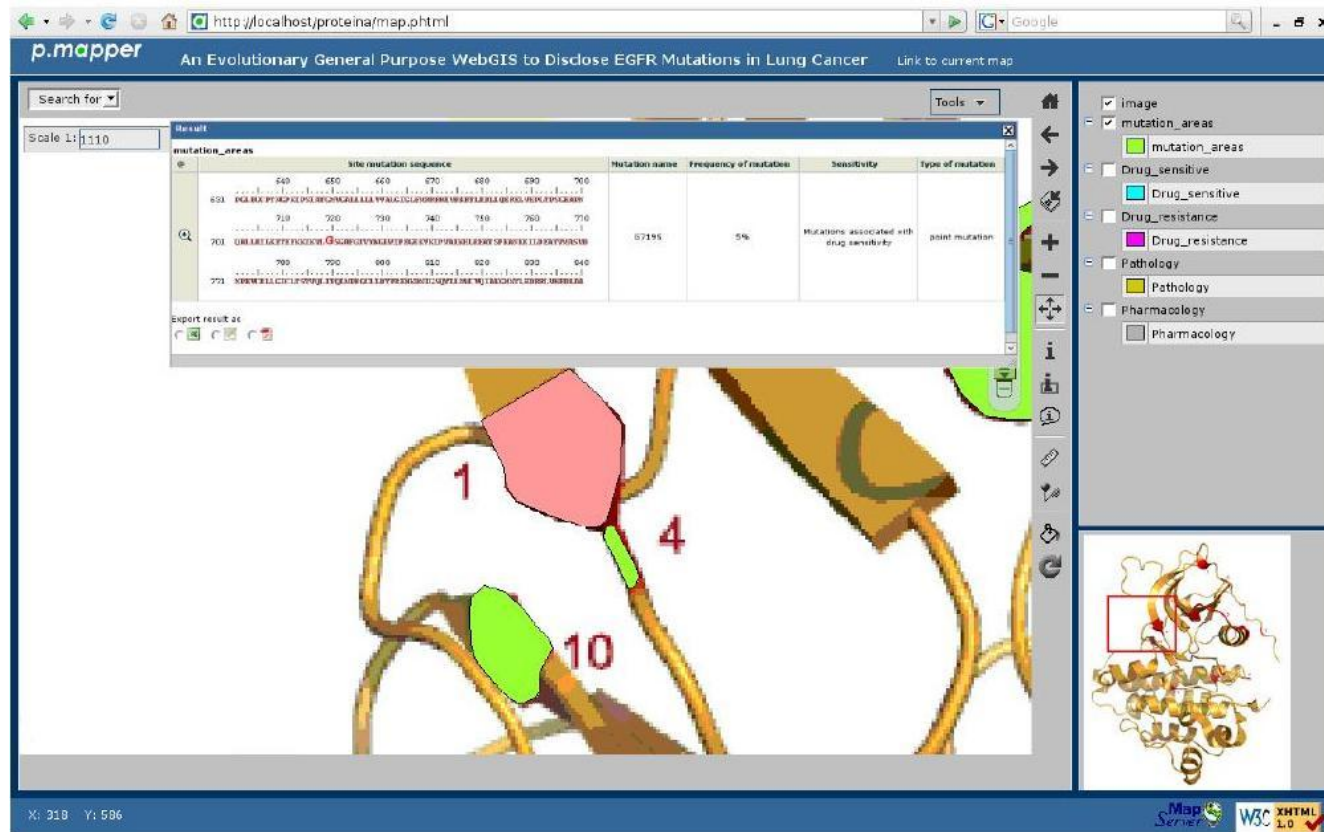
Among them there are:

the zoom and pan interface (without frames); by using them the user can browse the protein image easily.

(User can also zoom/pan through the map via keyboard keys, mouse wheel, reference map and slider bar).

Users can query data in the database with three functions: identify, select and search; in fact there is a fully featured attribute search, including suggest, select boxes, etc., and query results can show joins from DB and hyperlinks (query results can be also exported).

Functionalities



Query layout results is really flexible because it is managed via JavaScript and can be shown with a pop-in identification by moving the mouse over the map, or with an i-frame into the map (auto-identify function).

Functionalities

All data in the database can be printed out as HTML and PDF file.

The application has a useful tool to measure distance and areas on the map.

All data are figured also with HTML legends and various display styles of legend and layers/TOC, thanks to a very flexible configuration of functions, behavior and layout.

The application has a multilingual user interface and predefined languages are: English, Italian, French and German.

User can also add points of interest in the window with labels on the map.

In the first prototype of the WebGIS we provided five layers, each one containing information about EGFR mutations.

The first one contains information about mutations; in particular, in the database are collected data regarding the site mutation sequences, and the position where they happen, the mutations name, the frequencies with which they occur, and their typology (punctual or not).

The second and third contain information about drug sensitivity/resistance;

The fourth and fifth about pathology deriving from mutations, diseases, pharmacology experimentations and so on.

Moreover, a lot of items regarding no-gis-data are provided. This category includes a dataset not depending on mutations, but regarding EGFR protein and cancer relationship.

“No-Gis data”

The screenshot displays the p.mapper web application interface. The browser address bar shows <http://localhost/proteina/map.phtml>. The application title is "An Evolutionary General Purpose WebGIS to Disclose EGFR Mutations in Lung Cancer".

The main interface includes a search bar, a scale indicator (Scale 1: 5020), and a 3D protein structure on the left. A central window displays a table of mutations and a network diagram of protein-protein interactions on the right. The network diagram shows nodes representing proteins and edges representing interactions.

Pos	EGFR2953 Code	Description	
1201	23010010201	181 - 182 - 183 - 184 - 185 - 186 - 187 - 188 - 189 - 190 - 191 - 192 - 193 - 194 - 195 - 196 - 197 - 198 - 199 - 200 - 201 - 202 - 203 - 204 - 205 - 206 - 207 - 208 - 209 - 210 - 211 - 212 - 213 - 214 - 215 - 216 - 217 - 218 - 219 - 220 - 221 - 222 - 223 - 224 - 225 - 226 - 227 - 228 - 229 - 230 - 231 - 232 - 233 - 234 - 235 - 236 - 237 - 238 - 239 - 240 - 241 - 242 - 243 - 244 - 245 - 246 - 247 - 248 - 249 - 250 - 251 - 252 - 253 - 254 - 255 - 256 - 257 - 258 - 259 - 260 - 261 - 262 - 263 - 264 - 265 - 266 - 267 - 268 - 269 - 270 - 271 - 272 - 273 - 274 - 275 - 276 - 277 - 278 - 279 - 280 - 281 - 282 - 283 - 284 - 285 - 286 - 287 - 288 - 289 - 290 - 291 - 292 - 293 - 294 - 295 - 296 - 297 - 298 - 299 - 300 - 301 - 302 - 303 - 304 - 305 - 306 - 307 - 308 - 309 - 310 - 311 - 312 - 313 - 314 - 315 - 316 - 317 - 318 - 319 - 320 - 321 - 322 - 323 - 324 - 325 - 326 - 327 - 328 - 329 - 330 - 331 - 332 - 333 - 334 - 335 - 336 - 337 - 338 - 339 - 340 - 341 - 342 - 343 - 344 - 345 - 346 - 347 - 348 - 349 - 350 - 351 - 352 - 353 - 354 - 355 - 356 - 357 - 358 - 359 - 360 - 361 - 362 - 363 - 364 - 365 - 366 - 367 - 368 - 369 - 370 - 371 - 372 - 373 - 374 - 375 - 376 - 377 - 378 - 379 - 380 - 381 - 382 - 383 - 384 - 385 - 386 - 387 - 388 - 389 - 390 - 391 - 392 - 393 - 394 - 395 - 396 - 397 - 398 - 399 - 400 - 401 - 402 - 403 - 404 - 405 - 406 - 407 - 408 - 409 - 410 - 411 - 412 - 413 - 414 - 415 - 416 - 417 - 418 - 419 - 420 - 421 - 422 - 423 - 424 - 425 - 426 - 427 - 428 - 429 - 430 - 431 - 432 - 433 - 434 - 435 - 436 - 437 - 438 - 439 - 440 - 441 - 442 - 443 - 444 - 445 - 446 - 447 - 448 - 449 - 450 - 451 - 452 - 453 - 454 - 455 - 456 - 457 - 458 - 459 - 460 - 461 - 462 - 463 - 464 - 465 - 466 - 467 - 468 - 469 - 470 - 471 - 472 - 473 - 474 - 475 - 476 - 477 - 478 - 479 - 480 - 481 - 482 - 483 - 484 - 485 - 486 - 487 - 488 - 489 - 490 - 491 - 492 - 493 - 494 - 495 - 496 - 497 - 498 - 499 - 500 - 501 - 502 - 503 - 504 - 505 - 506 - 507 - 508 - 509 - 510 - 511 - 512 - 513 - 514 - 515 - 516 - 517 - 518 - 519 - 520 - 521 - 522 - 523 - 524 - 525 - 526 - 527 - 528 - 529 - 530 - 531 - 532 - 533 - 534 - 535 - 536 - 537 - 538 - 539 - 540 - 541 - 542 - 543 - 544 - 545 - 546 - 547 - 548 - 549 - 550 - 551 - 552 - 553 - 554 - 555 - 556 - 557 - 558 - 559 - 560 - 561 - 562 - 563 - 564 - 565 - 566 - 567 - 568 - 569 - 570 - 571 - 572 - 573 - 574 - 575 - 576 - 577 - 578 - 579 - 580 - 581 - 582 - 583 - 584 - 585 - 586 - 587 - 588 - 589 - 590 - 591 - 592 - 593 - 594 - 595 - 596 - 597 - 598 - 599 - 600 - 601 - 602 - 603 - 604 - 605 - 606 - 607 - 608 - 609 - 610 - 611 - 612 - 613 - 614 - 615 - 616 - 617 - 618 - 619 - 620 - 621 - 622 - 623 - 624 - 625 - 626 - 627 - 628 - 629 - 630 - 631 - 632 - 633 - 634 - 635 - 636 - 637 - 638 - 639 - 640 - 641 - 642 - 643 - 644 - 645 - 646 - 647 - 648 - 649 - 650 - 651 - 652 - 653 - 654 - 655 - 656 - 657 - 658 - 659 - 660 - 661 - 662 - 663 - 664 - 665 - 666 - 667 - 668 - 669 - 670 - 671 - 672 - 673 - 674 - 675 - 676 - 677 - 678 - 679 - 680 - 681 - 682 - 683 - 684 - 685 - 686 - 687 - 688 - 689 - 690 - 691 - 692 - 693 - 694 - 695 - 696 - 697 - 698 - 699 - 700 - 701 - 702 - 703 - 704 - 705 - 706 - 707 - 708 - 709 - 710 - 711 - 712 - 713 - 714 - 715 - 716 - 717 - 718 - 719 - 720 - 721 - 722 - 723 - 724 - 725 - 726 - 727 - 728 - 729 - 730 - 731 - 732 - 733 - 734 - 735 - 736 - 737 - 738 - 739 - 740 - 741 - 742 - 743 - 744 - 745 - 746 - 747 - 748 - 749 - 750 - 751 - 752 - 753 - 754 - 755 - 756 - 757 - 758 - 759 - 760 - 761 - 762 - 763 - 764 - 765 - 766 - 767 - 768 - 769 - 770 - 771 - 772 - 773 - 774 - 775 - 776 - 777 - 778 - 779 - 780 - 781 - 782 - 783 - 784 - 785 - 786 - 787 - 788 - 789 - 790 - 791 - 792 - 793 - 794 - 795 - 796 - 797 - 798 - 799 - 800 - 801 - 802 - 803 - 804 - 805 - 806 - 807 - 808 - 809 - 810 - 811 - 812 - 813 - 814 - 815 - 816 - 817 - 818 - 819 - 820 - 821 - 822 - 823 - 824 - 825 - 826 - 827 - 828 - 829 - 830 - 831 - 832 - 833 - 834 - 835 - 836 - 837 - 838 - 839 - 840 - 841 - 842 - 843 - 844 - 845 - 846 - 847 - 848 - 849 - 850 - 851 - 852 - 853 - 854 - 855 - 856 - 857 - 858 - 859 - 860 - 861 - 862 - 863 - 864 - 865 - 866 - 867 - 868 - 869 - 870 - 871 - 872 - 873 - 874 - 875 - 876 - 877 - 878 - 879 - 880 - 881 - 882 - 883 - 884 - 885 - 886 - 887 - 888 - 889 - 890 - 891 - 892 - 893 - 894 - 895 - 896 - 897 - 898 - 899 - 900 - 901 - 902 - 903 - 904 - 905 - 906 - 907 - 908 - 909 - 910 - 911 - 912 - 913 - 914 - 915 - 916 - 917 - 918 - 919 - 920 - 921 - 922 - 923 - 924 - 925 - 926 - 927 - 928 - 929 - 930 - 931 - 932 - 933 - 934 - 935 - 936 - 937 - 938 - 939 - 940 - 941 - 942 - 943 - 944 - 945 - 946 - 947 - 948 - 949 - 950 - 951 - 952 - 953 - 954 - 955 - 956 - 957 - 958 - 959 - 960 - 961 - 962 - 963 - 964 - 965 - 966 - 967 - 968 - 969 - 970 - 971 - 972 - 973 - 974 - 975 - 976 - 977 - 978 - 979 - 980 - 981 - 982 - 983 - 984 - 985 - 986 - 987 - 988 - 989 - 990 - 991 - 992 - 993 - 994 - 995 - 996 - 997 - 998 - 999 - 1000	

The network diagram shows nodes representing proteins and edges representing interactions. The nodes are labeled with protein names such as EGFR, ERBB2, ERBB3, ERBB4, and others. The edges represent interactions between these proteins.

The right sidebar contains a legend with the following items:

- image
- mutation_areas
 - mutation_areas
- Drug_sensitive
 - Drug_sensitive
- Drug_resistance
 - Drug_resistance
- Pathology
 - Pathology
- Pharmacology
 - Pharmacology

The bottom status bar shows coordinates X: 428 Y: 889 and logos for Map Server, W3C, and XHTML 1.0.

An example, of no gis data are the *protein2protein* interactions, that are viewable by clicking on the *i* icon of the toolbar

Conclusion

The strength of this system is to provide to researchers a complete and multidisciplinary easy-to-use service, that is usable from different places via Internet and can be developed with low costs and little maintenance work.

It is able to improve medical research in cancer mapping, analysing and exchanging data concerning proteins.

Potential users of our application include biologists, doctors, researchers, as well as private collectors of biomedical data/objects, and non-expert people that will obtain information regarding diseases, statistics, therapies, and so on.

Conclusion

This is possible thanks to the powerful tools of a typical WebGIS, such as WMS (Web Map Service), WFS (Web Feature Service), and WMC (Web Map Context). Using them it is possible to improve interoperability between applications, by creating some common interchange languages through common standards, so that applications can share and request vectors and rasters maps.

They can be produced in plain image format, through WMS; in vector map data and attributes (GML format), through WFS; and in a XML file format, saving WMS views through WMC.

In particular biomedical data, which often consist in different kinds of information, require simple instruments to be better arranged and displayed, so to ensure visibility, and to facilitate learning, research, therapy and so on.

Future Works

Next step will be to implement a 3-Dimensional WebGIS, where 3D space will be used to store biomedical images linked from different databases, with different items, providing data-mining, graphical queries, and so on.

This distributed 3D WebGIS could represent a huge improvement in biomedical science.

Moreover, our goal will be to put in relation biological functions, associated diseases, related genes and gene-to-gene characteristics, allowing to manage, visualize and interact various level of information; in order to improve interoperability between different teams, working worldwide on medical research in cancer.