A The SURFCOMP Program Suite

The SURFCOMP program suite consists of about 30.000 lines of code and performs the heuristic filtering process described in Figure 3-1 together with the preparations of the surfaces and the analysis of the search results. The main program is `surfcomp` which calculates the surface similarities and the functions for the consensus scoring (3.10). The generation and preparation of the molecular surfaces can be performed either via Sybyl 6.91 [2] and MOLCAD [24] or via the MSMS program [114] and a property calculator written in C++. All binary programs developed in this project are available for Linux via source archives or package files in the RPM package manager file format [126]. The evaluation of the results, the ranking, visualization and generation of alignment data, is done by the graphical user interface `surfcomp-monitor` which acts as a plug-in for the geometry viewer Geomview [1].

The input data (the molecular surface objects and data files) and the fundamental comparison parameters are stored and managed by a local or remote MySQL database and scripts are provided that automatically setup and fill the experiment-databases. For the preparation of the surfaces by Sybyl and MOLCAD a suite of SPL scripts is available that provides convenient tools especially for the setup of protein active site comparisons. The suite can handle a series of other surface file formats and is able to calculate basic surface properties like canonical curvatures [141] and the electrostatic potential (section 2.2.1) via the auxiliary program `propgenerator`.

The calls to `surfcomp` are usually invoked via a shell script which is generated from a template by the script `preparesurfcomp`. One can take any user-defined template for that script and it is thereby possible to distribute the single jobs by a scheduler to a Linux cluster or to other high performance computer systems.

A.1 Requirements

The following additional libraries are required

- the Linux operating system
- the Xerces-c XML library (version 2.1 or higher) [124]
- the MySQL client libraries (versions 3.23 or higher)
- Geomview (version 1.8 or higher)
- Perl (version 5.6 or higher) [125]
- A MySQL database server (versions 3.23 or higher) with read/write access for the user.

A.2 Availability

The source code is published under the Novartis open-source license and is available from the web [64] or the attached CD-ROM together with the binary packages for various Linux distributions. Binary executables for other platforms must be created from the source code. Especially compilation for other UNIX compliant systems should be possible with the provided installation tools.

B Publications

J. Chem. Inf. Comput. Sci., **44** (3), 837 -847, 2004.

SURFCOMP: A Novel Graph-Based Approach to Molecular Surface Comparison

Christian Hofbauer, Hans Lohninger, and András Aszódi

Novartis Institutes for BioMedical Research, Brunnerstrasse 59, A-1235 Vienna, Austria, and Institut für Chemische Technologien und Analytik, Technische Universität Wien, Getreidemarkt 9/151, A-1060 Vienna, Austria

Received October 28, 2003

Abstract. Analysis of the distributions of physicochemical properties mapped onto molecular surfaces can highlight important similarities or differences between compound classes, contributing to rational drug design efforts. Here we present an approach that uses maximal common subgraph comparison and harmonic shape image matching to detect locally similar regions between two molecular surfaces augmented with properties such as the electrostatic potential or lipophilicity. The complexity of the problem is reduced by a set of filters that implement various geometric and physicochemical heuristics. The approach was tested on dihydrofolate reductase and thermolysin inhibitors and was shown to recover the correct alignments of the compounds bound in the active sites.

Theory and Applications of Computational Chemistry, Gyeongju, Korea,
14. - 20. Feb. 2004

Molecular surface comparison with SURFCOMP: A novel graph-based approach

Christian Hofbauer, Hans Lohninger, and András Aszódi

Institut für Chemische Technologien und Analytik, Technische Universität Wien, Getreidemarkt 9/151, A-1060 Vienna, AUSTRIA
Novartis Institutes for BioMedical Research, Brunnerstrasse 59, A-1235 Vienna, AUSTRIA

Abstract. Analysis of the distributions of physicochemical properties mapped onto molecular surfaces can highlight important similarities or differences between compound classes, contributing to rational drug design efforts [131]. We have developed a method that uses a combination of graph theory, computer vision and computational chemistry to detect local surface similarities between small and medium sized molecules. Our approach is based on 3D structure search where maximal common subgraph isomorphism is used to detect local similarities between the pharmacophoric feature points of different molecules [91]. The extension of this principle to molecular surfaces is cumbersome, because treatment of the complete set of surface points instead of just a few feature points with NP-hard graph algorithms is not feasible. In order to

perform a reliable and fast detection of local surface similarities it is necessary to reduce the complexity of the problem by a set of filters that implement various geometric and physicochemical heuristics.

To achieve this we first generate a simplified representation of the surfaces consisting only of a set of critical points (corresponding to “hills” and “valleys” on the surface), augmented by their surrounding surface patches. Among all possible point pairs we first select those that show sufficient chemical similarity, judged by means of a fuzzy dissimilarity index [48] between physicochemical properties mapped onto the surface points. Then the curvature patterns around all remaining point pairs are compared by harmonic shape image matching [145] to discard points that are not embedded in a similar shape. Finally the distances and angles between combinations of similar pairs are checked to be within certain bounds to form an association graph that is simple enough for the clique detection. The cliques represent the local surface similarities and an alignment between the two molecular surfaces can be calculated based on the corresponding points. Finally the alignments can be clustered to reveal a picture of the total surface similarity between the two molecules.

We tested our method with a dataset of eight thermolysin inhibitors and recovered the correct alignments of the compounds bound in the active sites. The results were in good agreement with another surface-based comparison carried out on the same dataset [37]. We are now directing our efforts to the comparison of protein/protein surfaces and the incorporation of conformational flexibility.

C Abbreviations

ALA	alanine
ARG	arginine
ASP	aspartic acid
ATP	adenosine triphosphate
Bppm	bis(para-phosphophenyl)methane
CP	critical point
CPU	central processing unit
CRK	proto-oncogene C
CYS	cysteine
DHFR	dihydrofolate reductase
DNA	deoxyribonucleic acid
EAT-2	ews/fli1 activated transcript 2
EC	Enzyme Commission
ESP	electrostatic potential
FOL	folic acid
GLU	glutamic acid
GLY	glycine
GRB2	growth factor receptor-bound protein 2
HF	Hartree Fock
HOMO	highest occupied molecular orbital
HSI	harmonic shape image
ILE	isoleucine
LP	lipophilic potential
LUMO	lowest unoccupied molecular orbital
LYS	lysine
MTX	methotrexate
NADP(H)	nicotinamide adenine dinucleotide phosphate
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser enhancement Spectroscopy
PDB	Protein Data Bank
PHE	phenylalanine
PTP1B	protein tyrosine phosphatase 1B
PTPases	protein tyrosine phosphatases
pTyr	phosphorylated tyrosine
QSAR	quantitative structure activity relationship
QSD	quadratic shape descriptors
RAM	random access memory
RMS	root mean square

RMSD	root mean square deviation
SAP	SLAM← associated protein
SH2	SRC homology 2
SLAM	**s**ignaling lymphocyte activation molecule
SPL	Sybyl programming language
STI	surface topology index
TLN	Thermolysin
TMP	trimethoprim
WRB	Br-WR99210 *(compound name)*

D Indices

D.1 Figure Index

D.2 Table Index

D.3 Chart Index

E References

(1) Geomview, *http://www.geomview.org/* , 1.2.2002.

(2) *SYBYL 6.9*; Tripos Inc.: St. Louis, MO, 2003

(3) POV-Ray the "persistence of vision" raytracer, *http://www.povray.org/* , 15.3.2004.

(4) United Devices, *http://www.ud.com/home.htm* , 2004.

(5) Abagyan, R.; Totrov, M. High-throughput docking for lead generation *Current Opinion in Chemical Biology* **5:4**, 375-382, 2001.

(6) Amann, A.; Bangov, I. P.; Brickmann, J.; Dumitrescu, D.-D.; Klir, G. J.; Mezey, P. G.; Mislow, K.; Rouvray, D. H.; Xu, J. *Fuzzy Logic in Chemistry*; Academic Press: San Diego, CA, 1997.

(7) Appleton, T. Combinatorial chemistry and HTS - feeding a voracious process *Drug Discovery Today* **4:9**, 398-400, 1999.

(8) Aszódi, A. RazorBack 2.0 - linear algebra library, *unpublished work*.

(9) Atkins, P. W.; Friedman, R. S.; Editors. *Molecular Quantum Mechanics, 3rd Edition*; 1996.

(10) Ausiello, G.; Cesareni, G.; Helmer-Citterich, M. ESCHER: a new docking procedure applied to the reconstruction of protein tertiary structure *Proteins: Structure, Function, and Genetics* **28:4**, 556-567, 1997.

(11) Bailey, R. R.; Srinath, M. Orthogonal Moment Features for Use With Parametric and Non-Parametric Classifiers *IEEE Trans.Pattern Anal.Mach.Intell.* **18**, 389-399, 1996.

(12) Barrow, H. G.; Burstall, R. M. Subgraph isomorphism, matching relational structures and maximal cliques *Inf.Process.Lett.* **4:4**, 83-84, 1976.

(13) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank *Nucleic Acids Res.* **28:1**, 235-242, 2000.

(14) Bhatia, A.; E.áWolf On the circle polynomials of Zernike and related orthogonal sets *Proc.Cambridge.Philosophical.Society.* **50**, 40-48, 1954.

(15) Bladon, P. A rapid method for comparing and matching the spherical parameter surfaces of molecules and other irregular objects *J.Mol.Graphics* **7**, 130-137, 1989.

(16) Blakley, R. L.; Cocco, L. Dismutation of dihydrofolate by dihydrofolate reductase *Biochemistry* **23:11**, 2377-2383, 1984.

(17) Blaney, F.; Flinn, P.; Phippen, R.; Wyatt, R. Molecular surface comparison: Application to drug design *J.Mol.Graphics* **11**, 98-105, 1993.

(18) Blaney, J. M.; Dixon, J. S. A good ligand is hard to find: automated docking methods *Perspect.Drug Discovery Des.* **1:2**, 301-319, 1993.

(19) Boland, M. V.; Markey, M. K.; Murphy, R. F. Automated Recognition of Patterns Characteristic of Subcellular Structures in Fluorescence Microscopy Images *Cytometry* **33**, 366-375, 1998.

(20) Bondi, A. van der Waals volumes and radii *J.Phys.Chem.* **68:3**, 441-451, 1964.

(21) Botfield, M. C.; Green, J. SH2 and SH3 domains: choreographers of multiple signaling pathways *Annual Reports in Medicinal Chemistry* **30**, 227-237, 1995.

(22) Bowen, J. P.; Allinger, N. L. Molecular mechanics: the art and science of parameterization *Rev.Comput.Chem.* **2**, 81-97, 1991.

(23) Brickmann, J.; Bertling, H.; Bussian, B. M.; Goetze, T.; Knoblauch, M.; Waldherr-Teschner, M. MOLCAD - interactive molecular computer graphics on high-performance computers *Tagungsber.- Vortragstag., Ges.Dtsch.Chem., Fachgruppe Chem.-Inf.* **3rd**, 93-111, 1987.

(24) Brickmann, Jürgen, Goetze, Thomas, Heiden, Wolfgang, Moeckel, Gerd, Reiling, Stephan, Vollhardt, Horst, and Zachmann, Carl Dieter Interactive visualization of molecular scenarios with MOLCAD/SYBYL. *Data Visualization Mol. Sci* **1995**,

(25) Brint, A. T.; Willett, P. Algorithms for the identification of three-dimensional maximal common substructures *J.Chem.Inf.Comp.Sci* **27:4**, 152-158, 1987.

(26) Bron, C.; Kerbosch, J. Algorithm 457 - Finding all cliques of an undirected graph *Commun.ACM* **16:9**, 575-577, 1973.

(27) Broughton, H. B. A method for including protein flexibility in protein-ligand docking: improving tools for database mining and virtual screening *Journal of Molecular Graphics & Modelling* **18:3**, 247-257, 2000.

(28) Chan, B.; Lanyi, A.; Song, H. K.; Griesbach, J.; Simarro-Grande, M.; Poy, F.; Howie, D.; Sumegi, J.; Terhorst, C.; Eck, M. J. SAP couples Fyn to SLAM immune receptors *Nat.Cell Biol.* **5:2**, 155-160, 2003.

(29) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins *J.Med.Chem.* **42:25**, 5100-5109, 1999.

(30) Chau, P. L.; Dean, P. M. Molecular recoginition: 3D surface structure comparison by gnomic projection *J.Mol.Graphics* **5:2**, 97-100, 1987.

(31) Chothia, C.; Janin, J. Principles of protein-protein recognition *Nature* **256:5520**, 705-708, 1975.

(32) Connolly, M. L. Analytical molecular surface calculation *J.Appl.Crystallogr.* **16**, 548-558, 1983.

(33) Connolly, M. L. Solvent-accessible surfaces of proteins and nucleic acids *Science* **221:4612**, 709-713, 1983.

(34) Connolly, M. L. Shape complementarity at the hemoglobin α_{1} β1 subunit interface *Biopolymers* **25:7**, 1229-1247, 1986.

(35) Connolly, M. L. Shape distributions of protein topography *Biopolymers* **32:9**, 1215-1236, 1992.

(36) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Mertz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J.Am.Chem.Soc.* **117**, 5179-5197, 1995.

(37) Cosgrove, D.; Bayada, D.; Johnson, A. A novel method of aligning molecules by local surface shape similarity *J.Comput.-Aided Mol.Des.* **14:6**, 573-591, 2000.

(38) Davies, J. F.; Delcamp, T. J.; Prendergast, N. J.; Ashford, V. A.; Freisheim, J. H.; Kraut, J. Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate *Biochemistry* **29:40**, 9467-9479, 1990.

(39) Desbrun, M., Meyer, M., der, P., and Barr, A. *Discrete differential-geometry operators in nD,***2000.**

(40) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model *J.Am.Chem.Soc.* **107:13**, 3902-3909, 1985.

(41) Dijkstra, E. A note on two problems in connexion with graphs. *Numerische Mathematik* **1**, 269-271, 1959.

(42) Diller, David J. Homology models, high throughput docking, and drug design. *Abstracts of Papers, 222nd ACS National Meeting, Chicago, IL, United States, August 26-30* **2001**,

(43) Diller, D. J.; Li, R. Kinases, Homology Models, and High Throughput Docking *J.Med.Chem.* **46:22**, 4638-4647, 2003.

(44) Eck, M., DeRose, T., Duchamp, T., Hoppe, H., Lounsbery, M., and Stuetzle, W. *Multiresolution analysis of arbitrary meshes*, University of Washington, Seattle, **1995.**

(45) Eck, M., DeRose, T., Duchamp, T., Hoppe, H., Lounsbery, M., and Stuetzle, W. *Multiresolution analysis of arbitrary meshes*, University of Washington, Seattle, **1995.**

(46) Eells, J.; Sampson, L. Harmonic mappings of Riemannian manifolds *Amer.J.Math.* **86**, 109-160, 1964.

(47) Entzeroth, M. Emerging trends in high-throughput screening *Current opinion in pharmacology* **3:5**, 522-529, 2003.

(48) Exner, T. E.; Keil, M.; Brickmann, J. Pattern recognition strategies for molecular surfaces. I. Pattern generation using fuzzy set theory *Journal of Computational Chemistry* **23:12**, 1176-1187, 2002.

(49) Exner, T. E.; Keil, M.; Brickmann, J. Pattern recognition strategies for molecular surfaces. II. Surface complementarity *Journal of Computational Chemistry* **23:12**, 1188-1197, 2002.

(50) Fastovski, O. High throughput docking: past, present, and future *PharmaChem* **1:7/8**, 18-21, 2002.

(51) Fischer, D.; Lin, S. L.; Wolfson, H. L.; Nussinov, R. A geometry-based suite of molecular docking processes *J.Mol.Biol.* **248:2**, 459-477, 1995.

(52) Forman, M. and Schaffer, P. *Amadeus*, The Saul Zaentz Company, 1984.

(53) Free Software Foundation, The GCC-GNU project, *http://gcc.gnu.org/* , 27.3.2004.

(54) Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships. I. Partition coefficients as a measure of hydrophobicity *J.Comp.Chem.* **7:4**, 565-577, 1986.

(55) Goldman, B. B.; Wipke, W. T. QSD quadratic shape descriptors. 2. Molecular docking using quadratic shape descriptors (QSDock) *Proteins: Structure, Function, and Genetics* **38:1**, 79-94, 2000.

(56) Goldman, B. B.; Wipke, W. T. Quadratic Shape Descriptors. 1. Rapid Superposition of Dissimilar Molecules Using Geometrically Invariant Surface Descriptors *J.Chem.Inf.Comp.Sci* **40:3**, 644-658, 2000.

(57) SPDB viewer, *http://www.expasy.org/spdbv/,* accessed on 2004.

(58) Guntert, P. Automated NMR protein structure calculation *Progress in nuclear magnetic resonance spectroscopy* **43:3-4**, 105, 2003.

(59) Heiden, W.; Moeckel, G.; Brickmann, J. A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces *J.Comput.-Aided Mol.Des.* **7:5**, 503-514, 1993.

(60) Heiden, W.; Brickmann, J. Segmentation of protein surfaces using fuzzy logic *J.Mol.Graphics* **12:2**, 106-115, 1994.

(61) Hekmat-Nejad, M.; Rathod, P. K. Plasmodium falciparum: kinetic interactions of WR99210 with pyrimethamine-sensitive and pyrimethamine-resistant dihydrofolate reductase *Exp.Parasitol.* **87:3**, 222-228, 1997.

(62) Hew, Patrick C. and Alder, Michael D. *Zernike or orthogonal Fourier-Mellon Moments for representing and recognising printed digits*, Department of Mathematics, The University of Western Australia, **1998.**

(63) Hoey, J.; Little, J. J. Representation and recognition of complex human motion; In *Conference on Computer Vision and Pattern Recognition*; 2000; pp 752-759.

(64) Hofbauer, C., The SURFCOMP program suite, *http://teachme.tuwien.ac.at/surfcomp* , 2004.

(65) Hofbauer, C.; Lohninger, H.; Aszódi, A. SURFCOMP: A novel graph-based approach to molecular surface comparison *J.Chem.Inf.Comp.Sci* **44:3**, 837-847, 2004.

(66) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics *J.Mol.Graphics* **14**, 33-38, 1996.

(67) Hunter, T. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling *Cell* **80:2**, 225-236, 1995.

(68) Hwang, P. M.; Li, C.; Morra, M.; Lillywhite, J.; Muhandiram, D. R.; Gertler, F.; Terhorst, C.; Kay, L. E.; Pawson, T.; Forman-Kay, J. D.; Li, S. C. A "three-pronged" binding mechanism for the SAP/SH2D1A SH2 domain: structural basis and relevance to the XLP syndrome *EMBO J.* **21:3**, 314-323, 2002.

(69) Inglese, J. Expanding the HTS paradigm *Drug Discovery Today* **7:18**, S105-S106, 2002.

(70) International Union of Biochemistry *Enzyme Nomenclature 1992*; Academic Press: New York, 1993.

(71) Karp, R. M. Reducibility among combinatorial problems; In *Complexity of Computer Computations*; Miller, R. E., Thatcher, J. W., eds. Plenum Press: New York, 1972; pp 85-103.

(72) Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. Molecular surface recognition: determination of geometric fit

between proteins and their ligands by correlation techniques *Proceedings of the National Academy of Sciences of the United States of America* **89:6**, 2195-2199, 1992.

(73) Keil, M. *Modellierung und Vorhersage von Strukturen biomolekularer Assoziate auf der Basis von statistischen Datenbankanalysen.* PhD Thesis, Technische Universität Darmstadt, Darmstadt, **2002**.

(74) Khotanzad, A.; Hong, Y. H. Invariant image recognition by Zernike moments *IEEE Trans.Pattern Anal.Mach.Intell.* **12:5**, 489-497, 1990.

(75) Koshland, D. E., Jr. Correlation of structure and function in enzyme action *Science* **142:3599**, 1533-1541, 1963.

(76) Koshland, D. E., Jr. Role of flexibility in the specificity, control and evolutiion of enzymes *FEBS letters* **62**, 47-52, 1976.

(77) Lamdan, Y. and Wolfson, H. J. Geometric Hashing: A General and Efficient Model Based Recognition Scheme. *International Conference on Computer Vision* **1988**, 213-229.

(78) Latour, S.; Roncagalli, R.; Chen, R.; Bakinowski, M.; Shi, X.; Schwartzberg, P. L.; Davidson, D.; Veillette, A. Binding of SAP SH2 domain to FynT SH3 domain reveals a novel mechanism of receptor signalling in immune regulation *Nat.Cell Biol.* **5:2**, 149-154, 2003.

(79) Leach, A. R. *Molecular Modelling. Principles and Applications.*; Prentice Hall: Upper Saddle River, NJ, USA, 2001.

(80) Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility *J.Mol.Biol.* **55**, 379-400, 1971.

(81) Leicester, S. E.; Finney, J. L.; Bywater, R. P. Description of molecular surface shape using Fourier descriptors *J.Mol.Graphics* **6:2**, 104-8, 100, 1988.

(82) Lemmen, C.; Lengauer, T. Time-efficient flexible superposition of medium-sized molecules *J.Comput.-Aided Mol.Des.* **11:4**, 357-368, 1997.

(83) Li, C.; Iosef, C.; Jia, C. Y.; Han, V. K.; Li, S. S. Dual functional roles for the X-linked lymphoproliferative syndrome gene product SAP/SH2D1A in signaling through the signaling lymphocyte activation molecule (SLAM) family of immune receptors *J.Biol.Chem.* **278:6**, 3852-3859, 2003.

(84) Li, R.; Sirawaraporn, R.; Chitnumsub, P.; Sirawaraporn, W.; Wooden, J.; Athappilly, F.; Turley, S.; Hol, W. G. Three-dimensional structure of m. tuberculosis dihydrofolate reductase reveals opportunities for the design of novel tuberculosis drugs *J.Mol.Biol.* **295:2**, 307-323, 2000.

(85) Li, S. C.; Gish, G.; Yang, D.; Coffey, A. J.; Forman-Kay, J. D.; Ernberg, I.; Kay, L. E.; Pawson, T. Novel mode of ligand binding by the SH2 domain of the human XLP disease gene product SAP/SH2D1A *Curr.Biol.* **9:23**, 1355-1362, 1999.

(86) Lin, S. L.; Nussinov, R.; Fischer, D.; Wolfson, H. J. Molecular surface representations by sparse critical points *Prot.Struct.Func.Gen.* **18:1**, 94-101, 1994.

(87) Little, J. J., Hoey, J., and Boyd, J. Characterizing 2D Flow Fields with Zernike, *unpublished work.*

(88) Maclennan, A. J.; Shaw, G. A yeast SH2 domain *Trends Biochem.Sci.* **18:12**, 464-465, 1993.

(89) Markey, M. K.; Boland, M. V.; Murphy, R. F. Toward Objective Selection of Representative Microscope Images *Biophys.J.* **76**, 2230-2237, 1999.

(90) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. The conformational parameter in drug design: the active analog approach *ACS Symposium Series* **112.**, 205-226, 1979.

(91) Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists *J.Comput.-Aided Mol.Des.* **7:1**, 83-102, 1993.

(92) Masek, B. B. Molecular surface comparisons. *Molecular Similarity in Drug Design* **1995**,

(93) McLachlan, A. Gene duplications in the structural evolution of chymotrypsin *J.Mol.Biol.* **128:1**, 49-79, 1979.

(94) McPherson, J. D.; Marra, M.; Hillier, L.; Waterston, R. H.; Chinwalla, A.; Wallis, J.; Sekhon, M.; Wylie, K.; Mardis, E. R.; Wilson, R. K.; Fulton, R.; Kucaba, T. A.; Wagner-McPherson, C.; Barbazuk, W. B.; Gregory, S. G.; Humphray, S. J.; French, L.; Evans, R. S.; Bethel, G.; Whittaker, A.; Holden, J. L.; McCann, O. T.; Dunham, A.; Soderlund, C.; Scott, C. E.; Bentley, D. R.; Schuler, G.; Chen, H. C.; Jang, W.; Green, E. D.; Idol, J. R.; Maduro, V. V.; Montgomery, K. T.; Lee, E.; Miller, A.; Emerling, S.; Kucherlapati; Gibbs, R.; Scherer, S.; Gorrell, J. H.; Sodergren, E.; Clerc-Blankenburg, K.; Tabor, P.; Naylor, S.; Garcia, D.; de Jong, P. J.; Catanese, J. J.; Nowak, N.; Osoegawa, K.; Qin, S.; Rowen, L.; Madan, A.; Dors, M.; Hood, L.; Trask, B.; Friedman, C.; Massa, H.; Cheung, V. G.; Kirsch, I. R.; Reid, T.; Yonescu, R.; Weissenbach, J.; Bruls, T.; Heilig, R.; Branscomb, E.; Olsen, A.; Doggett, N.; Cheng, J. F.; Hawkins, T.; Myers, R. M.; Shang, J.; Ramirez, L.; Schmutz, J.; Velasquez, O.; Dixon, K.; Stone, N. E.; Cox, D. R.; Haussler, D.; Kent, W. J.; Furey, T.; Rogic, S.; Kennedy, S.; Jones, S.; Rosenthal, A.; Wen, G.; Schilhabel, M.; Gloeckner, G.; Nyakatura, G.; Siebert, R.; Schlegelberger, B.; Korenberg, J.; Chen, X. N.; Fujiyama, A.; Hattori, M.; Toyoda, A.; Yada, T.; Park, H. S.; Sakaki, Y.; Shimizu, N.; Asakawa, S.; Kawasaki, K.; Sasaki, T.; Shintani, A.; Shimizu, A.; Shibuya, K.; Kudoh, J.; Minoshima, S.; Ramser, J.; Seranski, P.; Hoff, C.; Poustka, A.; Reinhardt, R.; Lehrach, H. A physical map of the human genome *Nature* **409:6822**, 934-941, 2001.

(95) Meza, M. B. Bead-based HTS applications in drug discovery *Drug Discovery Today* **5:Supplement 1**, 38-41, 2000.

(96) Mezey, P. G. The degree of similarity of three-dimensional bodies: application to molecular shape analysis *J.Math.Chem.* **7:1**, 39-49, 1991.

(97) MichaelLounsbery, J. W. Multiresolution analysis for surface of arbitrary topological type *ACM.Transactions.on.Graphics.* **16**, 34-73, 1997.

(98) Mirau, P. A.; Heffner, S. A.; Bovey, F. A. Three-dimensional nuclear Overhauser effect/J-resolved spectroscopy *Journal of Magnetic Resonance (1969-1992)* **89:3**, 572-577, 1990.

(99) Morra, M. Structural basis for the interaction of the free SH2 domain EAT-2 with SLAM receptors in hematopoietic cells *The EMBO journal* **20:21**, 5840, 2001.

(100) Morra, M.; Howie, D.; Grande, M. S.; Sayos, J.; Wang, N.; Wu, C.; Engel, P.; Terhorst, C. X-linked lymphoproliferative disease: a progressive immunodeficiency *Annu.Rev.Immunol.* **19**, 657-682, 2001.

(101) Mukundan, R.; Ramakrishnan, K. R. *Moment functions in image analysis*; World Scientific: Singapore, 1998.

(102) Norel, R.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Shape complementarity at protein-protein interfaces *Biopolymers* **34:7**, 933-940, 1994.

(103) Norel, R.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. Molecular surface complementarity at protein-protein interfaces: the critical role played by surface normals at well placed, sparse, points in docking *J.Mol.Biol.* **252:2**, 263-273, 1995.

(104) Perutz, M. F. Structure of hemoglobin *Brookhaven.Symp.Biol.* **13**, 165-183, 1960.

(105) Polanski, J.; Gasteiger, J.; Wagener, M.; Sadowski, J. The comparison of molecular surfaces by neural networks and its applications to quantitative structure activity studies *Quant.Struct.-Act.Relat.* **17:1**, 27-36, 1998.

(106) Polshakov, V. I.; Morgan, W. D.; Birdsall, B.; Feeney, J. Validation of a new restraint docking method for solution structure determinations of protein-ligand complexes *Journal of Biomolecular NMR* **14:2**, 115-122, 1999.

(107) Poy, F.; Yaffe, M. B.; Sayos, J.; Saxena, K.; Morra, M.; Sumegi, J.; Cantley, L. C.; Terhorst, C.; Eck, M. J. Crystal structures of the XLP protein SAP reveal a class of SH2 domains with extended, phosphotyrosine-independent sequence recognition *Mol.Cell* **4:4**, 555-561, 1999.

(108) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. LU Decomposition and Its Applications.; In *Numerical Recipes in FORTRAN: The Art of Scientific Computing*; Cambridge University Press: Cambridge, England, 1992; pp 34-42.

(109) Puius, Y. A.; Zhao, Y.; Sullivan, M.; Lawrence, D. S.; Almo, S. C.; Zhang, Z. Y. Identification of a second aryl phosphate-binding site in protein-tyrosine phosphatase 1B: a paradigm for inhibitor design *Proc.Natl.Acad.Sci.U.S.A* **94:25**, 13420-13425, 1997.

(110) Raevsky, O. A.; Grigor'ev, V. Yu.; Kireev, D. B.; Zefirov, N. S. Complete Thermodynamic Description of H-Bonding in the Framework of Multiplicative Approach *Quant.Struct.-Act.Relat.* **11**, 49-63, 1992.

(111) Raevsky, O. A.; Skvortsov, V. S. 3D hydrogen bond thermodynamics (HYBOT) potentials in molecular modelling *J.Comput.-Aided Mol.Des.* **16:1**, 1-10, 2002.

(112) Rarey, M.; Wefing, S.; Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins *J.Comput.-Aided Mol.Des.* **10:1**, 41-54, 1996.

(113) Ritchie, D. W. A. Protein docking using spherical polar Fourier correlations *Prot.Struct.Func.Gen.* **39**, 178-194, 2000.

(114) Sanner, M. F., Python based software developments, *http://www.scripps.edu/~sanner/python/index.html* , 14.1.2004.

(115) Sanner, M. F., Olson, A. J., and Spehner, J. Fast and robust computation of molecular surfaces. *Proc. 11th ACM Symp. Comp. Geom* **1995**, C6-C7.

(116) Saunders, M. Stochastic exploration of molecular mechanics energy surfaces. Hunting for the global minimum *J.Am.Chem.Soc.* **109:10**, 3150-3152, 1987.

(117) Sawyer, T. K. Src homology-2 domains: structure, mechanisms, and drug discovery *Biopolymers* **47:3**, 243-261, 1998.

(118) Schaffhausen, B. SH2 domain structure and function *Biochim.Biophys.Acta* **1242:1**, 61-75, 1995.

(119) Schrödinger Inc., Schrödinger: Jaguar Program, *http://www.schrodinger.com/Products/jaguar.html* , 29.3.2004.

(120) Schweighoffer, T. Investigating the Phosphatase Activities of SH2 Domains, *unpublished work*.

(121) Schweitzer, B. I.; Dicker, A. P.; Bertino, J. R. Dihydrofolate reductase as a therapeutic target *FASEB J.* **4:8**, 2441-2452, 1990.

(122) Shoichet, B. K.; Kuntz, I. D. Matching chemistry and shape in molecular docking *Protein Eng.* **6:7**, 723-732, 1993.

(123) Siek, J., Lee, L.-Q., and Lumsdaine, A., The Boost graph library, *http://www.boost.org/libs/graph/doc/index.html* , 2004.

(124) The Apache Software Foundation, Xerces C++ Parser, *http://xml.apache.org/xerces-c/index.html* , 2004.

(125) The Perl Foundation, The Perl Directory, *http://www.perl.org/* , 2004.

(126) the RPM community, RPM Package Manager, *http://www.rpm.org/* , 2004.

(127) van der Waals, J. D. *Die Zustandsgleichung*;

(128) van der Waals, J. D. The volume of the molecule and the volume of the constituent atoms *Verslag van de Gewone Vergadering van de Afdeling Natuurkunde, Koninklijke Nederlandse Akademie van Wetenschappen* **22**, 782-792, 1914.

(129) Vangrevelinghe, E.; Zimmermann, K.; Schoepfer, J.; Portmann, R.; Fabbro, D.; Furet, P. Discovery of a Potent and Selective Protein Kinase CK2 Inhibitor by High-Throughput Docking *J.Med.Chem.* **46:13**, 2656-2662, 2003.

(130) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A.; Gocayne, J. D.; Amanatides, P.; Ballew, R. M.; Huson, D. H.; Wortman, J. R.; Zhang, Q.; Kodira, C. D.; Zheng, X. H.; Chen, L.; Skupski, M.; Subramanian, G.; Thomas, P. D.; Zhang, J.; Gabor Miklos, G. L.; Nelson, C.; Broder, S.; Clark, A. G.; Nadeau, J.; McKusick, V. A.; Zinder, N.; Levine, A. J.; Roberts, R. J.; Simon, M.; Slayman, C.; Hunkapiller, M.; Bolanos, R.; Delcher, A.; Dew, I.; Fasulo, D.; Flanigan, M.; Florea, L.; Halpern, A.; Hannenhalli, S.; Kravitz, S.; Levy, S.; Mobarry, C.; Reinert, K.; Remington, K.; Abu-Threideh, J.; Beasley, E.; Biddick, K.; Bonazzi, V.; Brandon, R.; Cargill, M.; Chandramouliswaran, I.; Charlab, R.; Chaturvedi, K.; Deng, Z.; Di, F., V; Dunn, P.; Eilbeck, K.; Evangelista, C.; Gabrielian, A. E.; Gan, W.; Ge, W.; Gong, F.; Gu, Z.; Guan, P.; Heiman, T. J.; Higgins, M. E.; Ji, R. R.; Ke, Z.; Ketchum, K. A.; Lai, Z.; Lei, Y.;

Li, Z.; Li, J.; Liang, Y.; Lin, X.; Lu, F.; Merkulov, G. V.; Milshina, N.; Moore, H. M.; Naik, A. K.; Narayan, V. A.; Neelam, B.; Nusskern, D.; Rusch, D. B.; Salzberg, S.; Shao, W.; Shue, B.; Sun, J.; Wang, Z.; Wang, A.; Wang, X.; Wang, J.; Wei, M.; Wides, R.; Xiao, C.; Yan, C.; Yao, A.; Ye, J.; Zhan, M.; Zhang, W.; Zhang, H.; Zhao, Q.; Zheng, L.; Zhong, F.; Zhong, W.; Zhu, S.; Zhao, S.; Gilbert, D.; Baumhueter, S.; Spier, G.; Carter, C.; Cravchik, A.; Woodage, T.; Ali, F.; An, H.; Awe, A.; Baldwin, D.; Baden, H.; Barnstead, M.; Barrow, I.; Beeson, K.; Busam, D.; Carver, A.; Center, A.; Cheng, M. L.; Curry, L.; Danaher, S.; Davenport, L.; Desilets, R.; Dietz, S.; Dodson, K.; Doup, L.; Ferriera, S.; Garg, N.; Gluecksmann, A.; Hart, B.; Haynes, J.; Haynes, C.; Heiner, C.; Hladun, S.; Hostin, D.; Houck, J.; Howland, T.; Ibegwam, C.; Johnson, J.; Kalush, F.; Kline, L.; Koduru, S.; Love, A.; Mann, F.; May, D.; McCawley, S.; McIntosh, T.; McMullen, I.; Moy, M.; Moy, L.; Murphy, B.; Nelson, K.; Pfannkoch, C.; Pratts, E.; Puri, V.; Qureshi, H.; Reardon, M.; Rodriguez, R.; Rogers, Y. H.; Romblad, D.; Ruhfel, B.; Scott, R.; Sitter, C.; Smallwood, M.; Stewart, E.; Strong, R.; Suh, E.; Thomas, R.; Tint, N. N.; Tse, S.; Vech, C.; Wang, G.; Wetter, J.; Williams, S.; Williams, M.; Windsor, S.; Winn-Deen, E.; Wolfe, K.; Zaveri, J.; Zaveri, K.; Abril, J. F.; Guigo, R.; Campbell, M. J.; Sjolander, K. V.; Karlak, B.; Kejariwal, A.; Mi, H.; Lazareva, B.; Hatton, T.; Narechania, A.; Diemer, K.; Muruganujan, A.; Guo, N.; Sato, S.; Bafna, V.; Istrail, S.; Lippert, R.; Schwartz, R.; Walenz, B.; Yooseph, S.; Allen, D.; Basu, A.; Baxendale, J.; Blick, L.; Caminha, M.; Carnes-Stine, J.; Caulk, P.; Chiang, Y. H.; Coyne, M.; Dahlke, C.; Mays, A.; Dombroski, M.; Donnelly, M.; Ely, D.; Esparham, S.; Fosler, C.; Gire, H.; Glanowski, S.; Glasser, K.; Glodek, A.; Gorokhov, M.; Graham, K.; Gropman, B.; Harris, M.; Heil, J.; Henderson, S.; Hoover, J.; Jennings, D.; Jordan, C.; Jordan, J.; Kasha, J.; Kagan, L.; Kraft, C.; Levitsky, A.; Lewis, M.; Liu, X.; Lopez, J.; Ma, D.; Majoros, W.; McDaniel, J.; Murphy, S.; Newman, M.; Nguyen, T.; Nguyen, N.; Nodell, M. The sequence of the human genome *Science* **291:5507**, 1304-1351, 2001.

(131) Via, A.; Ferrè, F.; Brannetti, B.; Helmer-Citterich, M. Protein surface similarities: A survey of methods to describe and compare protein surfaces *Cell.Mol.Life Sci.* **57**, 1970-1977, 2000.

(132) Wagener, M.; Sadowski, J.; Gasteiger, J. Autocorrelation of molecular surface properties for modeling corticosteroid binding globulin and cytosolic Ah receptor activity by neural networks *J.Am.Chem.Soc.* **117:29**, 7769-7775, 1995.

(133) Wang, H. Grid-search molecular accessible surface algorithm for solving the protein docking problem *Journal of Computational Chemistry* **12:6**, 746-750, 1991.

(134) Wang, J.; Hou, T.; Chen, L.; Xu, X. Automated docking of peptides and proteins by genetic algorithm *Chemometrics and Intelligent Laboratory Systems* **45:1,2**, 281-286, 1999.

(135) Wang, R.; Wang, S. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment *J.Chem.Inf.Comp.Sci* **41:5**, 1422, 2001.

(136) Watson, J. D.; Crick, F. H. The structure of DNA *Cold Spring Harb.Symp.Quant.Biol.* **18**, 123-131, 1953.

(137) Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites *Chemistry & Biology* **3:6**, 449-462, 1996.

(138) Whitley, D. C. Van der Waals surface graphs and molecular shape *J.Math.Chem.* **23**, 377-397, 1998.

(139) Wüthrich, K. Three-dimensional structures of noncrystalline proteins observed by nuclear magnetic resonance *Chemica Scripta* **29A**, 23-26, 1989.

(140) Yang, J. M.; Kao, C. Y. Flexible ligand docking using a robust evolutionary algorithm *Journal of Computational Chemistry* **21:11**, 988-998, 2000.

(141) Zachmann, C. D.; Heiden, W.; Schlenkrich, M.; Brickmann, J. Topological analysis of complex molecular surfaces *J.Comp.Chem.* **13:1**, 76-84, 1992.

(142) Zadeh, L. A. Fuzzy Sets *Inform.Control.* **8**, 338-353, 1965.

(143) Zernike, F. Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode *Physica.* **1**, 689-704, 1934.

(144) Zhang, D. *Harmonic Shape Images: A 3D free-form surface representation and its applications in surface matching.* PhD Thesis, Carnegie Mellon University, Pittsburgh, **1999**.

(145) Zhang, D. and Herbert, M. Harmonic Maps and their applications in surface matching. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '99)* **1999**,