

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Structure-based design of an inhibitor of the conformational change of influenza hemagglutinin

Permalink

<https://escholarship.org/uc/item/0h5400sw>

Author

Bodian, Dale Lesley

Publication Date

1992

Peer reviewed|Thesis/dissertation

STRUCTURE-BASED DESIGN OF AN INHIBITOR OF THE
CONFORMATIONAL CHANGE OF INFLUENZA HEMAGGLUTININ
by

DALE LESLEY BODIAN

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOCHEMISTRY

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



copyright 1992
by
Dale Lesley Bodian

Preface

Any multidisciplinary endeavor requires contributions from colleagues specializing in diverse fields. This project would have been impossible without the collaborative atmosphere at UCSF fostering research at the boundaries of disciplines and the openness of my advisors, Tack Kuntz and Judy White, toward fields distant from their own.

Thanks are due to the past and present members of both the White and Kuntz laboratories for helpful discussions, for explaining new techniques, for sharing reagents and computer programs, and for maintaining computers and laboratory equipment. In particular, Renee DesJarlais was invaluable for her insightful critiques of oral presentations and for teaching me about DOCK, UNIX, and MIDAS. Brian Shoichet and Elaine Meng wrote versions 2 and 3 of DOCK and incorporated electrostatics into the programs. George Kemble taught me the biological experiments and did much of the work on the 212C, 216C mutant of hemagglutinin. Diane Mason, Anne Bartfay, and Manjita Bhaumik provided technical assistance. Eric Pettersen and other members of the UCSF Computer Graphics Laboratory gave patient help in hardware and software support. Patricia Caldera and Norm Oppenheimer, both in the Pharmaceutical Chemistry Department of UCSF, contributed expertise in organic synthesis.

Several groups outside of UCSF also contributed to this project. The members of SERES laboratories, Jay Stearns, Joe Nugent, Bryan

Yamasaki, and Dick Buswell played roles in designing, purifying, characterizing, and synthesizing compounds. The Influenza Group at the Centers for Disease Control in Atlanta kindly provided facilities and assistance for growing huge quantities of virus. The crystallographic work has been begun by Mike Eisen in Don Wiley's laboratory at Harvard University.

Thanks are due to my family for not giving me too hard a time about moving to California, for coming to visit, and for expressing interest in my work.

Special thanks go to Judy White who found space in her lab (the most valuable commodity at UCSF) for someone with minimal experimental experience and a proposal for a project with only a small chance of success. The most extra special thanks of all go to my advisor, Tack Kuntz, for support in every aspect of my graduate career: for allowing me to keep both a SUN workstation and an IRIS on my desk, for sending me to international meetings and to Harvard and Atlanta, for helping me debug experiments, for allowing me the freedom to choose a difficult project, for encouraging the flow of ideas between biologists, chemists, and theoreticians, and for teaching me the science and the politics required for productive research.

Abstract

Fusion between host cell and viral membranes is a crucial early step in the reproduction of enveloped animal viruses. In influenza viruses this fusion activity is mediated by the hemagglutinin protein. As a prerequisite to fusion, hemagglutinin undergoes an irreversible conformational change. A computer-aided inhibitor design method, DOCK, was used to identify a site on hemagglutinin that might regulate the fusion-inducing conformational change and to select ligands that might bind to the site. A family of small quinones and hydroquinones was found to inhibit the conformational change. The most active analog to date has an IC_{50} between 1 and 10 μM . Representative compounds inhibited hemagglutinin-mediated hemolysis, influenza-induced syncytia formation, and viral infectivity. These inhibitors can serve as lead compounds for anti-influenza drug design and as probes of the mechanism of the conformational change. The strategy of inhibiting infectivity by preventing the fusion-inducing conformational change is applicable to all enveloped viruses.

Structure-based inhibitor design methods such as DOCK identify ligands with steric and chemical properties complementary to those of a receptor site. However, drugs must also be specific for their intended binding site. The program DIFFDOCK is a modification of DOCK that begins to incorporate this concept into the inhibitor design process. Orientations of candidate ligands are generated over the entire target site, but only those configurations docked to specified subsites are evaluated further. By focusing orientation space to regions differentiating two homologous enzymes, DIFFDOCK can be used to identify selective inhibitors.

Contents

List of Tables	vii
List of Figures	ix
Chapter 1. Introduction	1
Chapter 2. Influenza Hemagglutinin:	
A Test Case for <i>De Novo</i> Drug Design	11
Chapter 3. Materials and Methods	19
Chapter 4. Site Selection and Inhibitor Design Round I	47
Chapter 5. Experimental Testing of Round I Compounds	84
Chapter 6. Inhibitor Design Round II	112
Chapter 7. Experimental Testing of Round II Compounds	130
Chapter 8. Discussion	171
Chapter 9. Further Studies on the Conformational Change of Influenza Hemagglutinin	186
Chapter 10. Differential Docking:	
Computer-Aided Design of Selective Inhibitors	209
Appendix A. Chemical Synthesis and Characterization	235
Appendix B. Hemagglutinin Sequences	244
Appendix C. Program Listings	265

List of Tables

Table 4-1.	Definition of sections for the initial molecular surface calculation.....	48
Table 4-2.	Parameters used in sphere generation.....	49
Table 4-3.	SPHGEN clusters.....	50
Table 4-4.	Residues comprising the tailla site.....	55
Table 4-5.	Conservation of the tailla site.....	56
Table 4-6.	DOCK1.1 parameter values.....	58
Table 4-7.	DOCK1.1 output summaries.....	60
Table 4-8.	Molecular skeletons from the CCD.....	61
Table 4-9.	Round I compounds.....	69
Table 4-10.	DOCK1.1 parameter values for docking to the interface site.....	79
Table 4-11.	DOCK1.1 output profile.....	79
Table 4-12.	Ligands selected from docking to the interface site.....	80
Table 5-1.	Effect of norakin in the SPA.....	87
Table 5-2.	Round I compounds in the IP.....	90
Table 5-3.	Round I compounds in the SPA.....	91
Table 5-4.	Round I compounds in the IP with other antibodies.....	92
Table 5-5.	Round I compounds and proteinase K digestions.....	93
Table 5-6.	Round I compounds and the aggregation assay.....	95
Table 5-7.	Round I compounds in hemolysis.....	97
Table 5-8.	Plaque assays with Round I compounds.....	99
Table 5-9.	Round I compounds in infectivity.....	101
Table 6-1.	Input to DISTMAP.....	114
Table 6-2.	Input to DELPHI.....	114
Table 6-3.	Input to DOCK2.1.....	115
Table 6-4.	DOCK2.1 output profiles.....	115
Table 6-5.	Selected FCD compounds.....	117
Table 6-6.	Input to CLUSTER.....	118
Table 6-7.	Input to DISTMAP2.....	119
Table 6-8.	Input to DOCK2.1 for high resolution runs.....	120

Table 6-9.	DOCK2.1 output profiles from high resolution runs.....	121
Table 6-10.	Selected FCD compounds from high resolution runs.....	122
Table 6-11.	Round II compounds.....	123
Table 7-1.	Round II compounds in the SPA.....	134
Table 7-2.	SPA dilution experiment with compound 83.....	136
Table 7-3.	Derivatives of 83 in the SPA.....	141
Table 7-4.	Derivatives of 83 in hemolysis.....	147
Table 7-5.	Derivatives of 83 on infectivity and cell viability.....	154
Table 7-6.	IC₅₀ values for the derivatives of 83.....	155
Table 7-7.	Structures of derivatives of compound 83.....	158
Table 7-8.	Input to DOCK3.0.....	166
Table 7-9.	Force field scores of the active compounds.....	166
Table 9-1.	IP with wt- and Cys-HA.....	195
Table 9-2.	Conformational changes as a function of temperature.....	198
Table 10-1.	Differential docking to M192A.....	220
Table 10-2.	Effect of varying the number of differential spheres.....	222
Table 10-3.	Docking LPF into chymotrypsin.....	225
Table A-1.	Purification and characterization of derivatives of 83.....	242
Table A-2.	Elemental analysis results.....	243

List of Figures

Figure 1-1.	Structure-based drug design cycle.....	3
Figure 2-1.	Initial steps of infection by an influenza virus.....	12
Figure 2-2.	α -Carbon backbone of BHA.....	15
Figure 4-1.	Location of the tail1A site on the BHA trimer.....	52
Figure 4-2.	Surface of the tail1A site contributed by the fusion peptide.....	53
Figure 4-3.	The tail1A site colored by chemical functionality.....	54
Figure 4-4.	Amantadine docked to the tail1A site.....	65
Figure 4-5.	Designed adamantyl derivatives.....	66
Figure 4-6.	3',5'-O-O-bis(carboxymethyl)-2'-deoxycytidine.....	67
Figure 5-1.	Norakin.....	85
Figure 5-2.	Norakin in hemolysis.....	86
Figure 5-3.	Norakin in the SPA as a function of pH.....	88
Figure 5-4.	Amantadine in hemolysis.....	98
Figure 7-1.	Comparison of IP and SPA.....	132
Figure 7-2.	Dose response of 83 in the SPA.....	135
Figure 7-3.	SPA with 83 as a function of time at low pH.....	138
Figure 7-4.	Structures of compounds 83K and 83A.....	139
Figure 7-5.	Compounds 83, 83K, 83A in the SPA.....	140
Figure 7-6.	Hemolysis with compound 91 as a function of pH.....	148
Figure 7-7.	Fusion from without.....	151
Figure 7-8.	Crystallographic difference density.....	164
Figure 7-9.	Comparison of IC ₅₀ and force field score.....	167
Figure 7-10.	Docked configuration of compound 90 superimposed on the crystallographic difference density.....	168
Figure 8-1.	Structural summary of the active compounds.....	172
Figure 9-1.	Antibody epitopes.....	187
Figure 9-2.	Location of the engineered disulfide bonds.....	188
Figure 9-3.	Two-stage model of the conformational change.....	189
Figure 9-4.	Time course of conformational changes of wt- and Cys-HA.....	197
Figure 9-5.	Time course of conformational changes at 0° and 37°C.....	198

Figure 9-6. Revised model of the conformational change.....	201
Figure 10-1. Active site of M192A.....	219
Figure 10-2. Specificity pocket of γ-chymotrypsin.....	224

Chapter 1

Introduction

Drug discovery ventures aim to find new bioactive compounds that are safer or more effective than existing chemotherapeutic agents or that treat diseases for which there are no known therapies. All methods require identification of a lead compound before improvements in activity can be made. Random screening trials [1, 2, 3] which test arbitrarily selected natural products or synthetic compounds for activity in a battery of biological assays and serendipity [4] have been the predominant means of discovering novel lead structures. A considerable drawback to these methods is that they depend entirely on chance and require screening of large numbers of chemicals. The probability of discovering an active agent is much increased by logically selecting chemicals for testing against specific targets [3]. Screening extracts of plants or herbs used for medicinal purposes by native peoples has somewhat mitigated the randomness of compound selection [3], but the number of targets is limited.

Rational design methods are based on drug-receptor theory [5]. Drugs are considered to exert their effects by binding to a particular receptor [6, 7]. Drug development, then, involves designing compounds that will “fit” into the binding site of the receptor. Generally this involves modification of known substrates or

inhibitors [8]. Methods such as QSAR (quantitative structure activity relationships) analyze data from sets of inhibitory compounds of various efficacies to predict the properties required for activity [9]. This method can efficiently optimize the activity of known inhibitors but has had limited success in discovering novel lead structures [10]. Design of transition state analogs has been successful in discovering inhibitors chemically distinct from the natural substrates of a reaction, but application is limited to receptors with known enzymatic activity.

A new method being developed is structure-based drug design [11]. Rather than inferring the properties of the receptor site from the properties of known ligands, this method uses the experimentally determined three-dimensional structure of a receptor in conjunction with computer graphics to predict functionalities conducive to binding. This strategy has an enormous advantage over earlier approaches in that it can propose structures not obviously related to known ligands. Its disadvantage of requiring a three-dimensional structure of the target site is becoming less significant as more structures are solved, and as new methods are devised for modeling receptor geometry [12, 13].

Structure-based drug design is an iterative process (Figure 1-1) [14]. Structural information about the binding mode of a lead compound is used to design inhibitors with greater complementarity to the site. The new compounds are synthesized and tested, and the structures of the active analogs bound to the site determined. The cycle is repeated until a potential drug possessing the desired properties is discovered. Structure determination is encouraged at

each iteration since small modifications of an inhibitor can lead to major and unexpected changes in its interaction with the receptor [15]. Current structure-based design methods are meant to identify *inhibitors* that can serve as lead compounds for drug design. These methods do not yet consider properties required of *drugs*, such as bioavailability and nontoxicity.

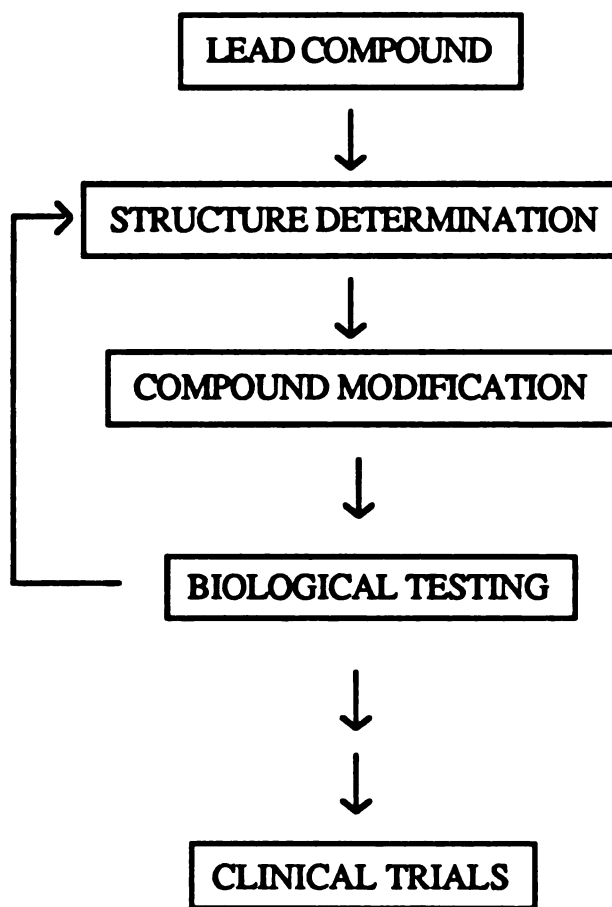


Figure 1-1. Structure-based drug design cycle.

Several novel inhibitors targeted against previously recognized sites have been developed by structure-based design methods [16, 17, 18, 19]. However, since no prior knowledge of substrates or inhibitors is required for structure-based design, newly discovered sites may be targeted. In theory, it should be possible to identify clefts or cavities in a receptor surface and design compounds *de novo* to bind to them. Such an approach has been successful in the design of covalent anti-sickling agents targeted to oxyhemoglobin [20]. Discovery of noncovalent inhibitors adds the difficulty of demanding a lead compound with sufficiently high affinity to be detected.

Several computer algorithms have been developed to aid in the design of novel inhibitors based on the structure of a receptor site [21, 22, 23, 24, 25]. One such algorithm is being developed at UCSF. DOCK [26, 27, 28] both identifies potential binding sites on a receptor and suggests structures of compounds that might bind to the selected site. It can reproduce crystallographic configurations of known ligand-receptor complexes [27, 29, 30], and has been used with partial success in design projects targeting enzyme active sites. For both HIV protease [31, 32] and thymidylate synthase [33], use of DOCK has led to discovery of novel inhibitors, although crystal structures of the enzyme-inhibitor complexes revealed binding modes or ligand geometries somewhat different from the predictions.

In a third test case, DOCK is being used to design inhibitors of the conformational change of influenza hemagglutinin. Unlike HIV protease and thymidylate synthase, hemagglutinin is not an enzyme and has no obvious active site. A new site had to be discovered that could prevent the conformational change of the receptor upon ligand

binding. Since a previously unrecognized site with no known substrates or inhibitors was targeted, this project tests the success of the current DOCK methodologies in developing a completely *de novo* inhibitor.

This thesis describes three projects. Parts of each project are contained in manuscripts submitted for publication. The hemagglutinin inhibitor design project is fully described in Chapters 2 - 8. An introduction to hemagglutinin and influenza viruses as well as the strategy undertaken to combat the conformational change are described in Chapter 2. Chapter 3 contains the experimental and computational methodologies used for the design and testing of the potential inhibitors. Discovery of a target site and the first round of design using DOCK1.1 [26] are the subject of Chapter 4. Chapter 6 describes the second round of modeling with DOCK2.1 [28]. The experimental results from testing the Round I and Round II compounds, both the successes and the failures, are given in Chapters 5 and 7, respectively. Chapter 8 summarizes the results and discusses their implications for DOCK in particular and rational drug design in general, for antiviral chemotherapy, and for the mechanism of the conformational change. Further biochemical studies regarding hemagglutinin's conformational change are described in Chapter 9.

In addition to high affinity, a good drug must have selectivity since binding to sites other than the targeted site could cause undesirable side effects [34]. The program DIFFDOCK begins to incorporate this concept into the DOCK method. A description of DIFFDOCK and some sample results are given in Chapter 10. The code itself is listed in Appendix C.

References

1. **Hamburger, M., Marston, A. and Hostettmann, K., "Search for new drugs of plant origin." *Advances in Drug Research* 20 167-209 (1991).**
2. **Che, C.-T., "Marine products as a source of antiviral drug leads." *Drug Development Research* 23 201-218 (1991).**
3. **Korolkovas, A. and Burckhalter, J. H., "Development of Drugs." In Essentials of Medicinal Chemistry. (John Wiley & Sons, New York, 1976) pp. 12-43.**
4. **Roberts, R. R., Serendipity. Accidental Discoveries in Science. (Wiley, New York, 1989).**
5. **Dean, P. M., Molecular Foundations of Drug-Receptor Interaction. (Cambridge University Press, Cambridge, 1987).**
6. **Ehrlich, P., "The relationship existing between chemical constitution, distribution and pharmacological action." In F. Himmelweit (Ed.), The Collected Papers of Paul Ehrlich. (Pergamon Press, London, 1956) pp. 596-618.**
7. **Langley, J. N., "On the physiology of salivary secretion II. On the mutual antagonism of atropin and pilocarpin, having special reference to their relations in the submaxillary gland of the cat." *J Physiol* 1 339-369 (1878).**
8. **Jack, D., "The challenge of drug discovery." *Drug Design and Delivery* 4 167-186 (1989).**

9. Martin, Y. C., "A practitioner's perspective of the role of quantitative structure-activity analysis in medicinal chemistry." *J Med Chem* **24** 229-237 (1981).
10. Austel, V., "Drug design: principles and techniques." In A. S. Horn and C. J. DeRanter (Eds.), Xray Crystallography and Drug Action. (Clarendon Press, Oxford, 1984) pp. 441-460.
11. Goodford, P. J., "Drug design by the method of receptor fit." *J Med Chem* **27** 557-564 (1984).
12. Folkers, G., Trumpff-Kallmeyer, S., Gutbrod, O., Krickl, S., Fetzner, J. and Keil, G. M., "Computer-aided active-site-directed modeling of the Herpes Simplex Virus 1 and human thymidine kinase." *Journal of Computer-Aided Molecular Design* **5** 385-404 (1991).
13. Cushman, D. W., Cheung, H. S., Sabo, E. F. and Ondetti, M. A., "Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids." *Biochemistry* **16** 5484-5491 (1977).
14. Hol, W. G. J., "Protein crystallography and computer graphics - toward rational drug design." *Angewandte Chemie* **25** 767-778 (1986).
15. Sauter, N. K., Glick, G. D., Crowther, R. L., Park, S.-J., Eisen, M. B., Skehel, J. J., Knowles, J. R. and Wiley, D. C., "Crystallographic detection of a second ligand binding site in influenza virus hemagglutinin." *PNAS* **89** 324-328 (1992).
16. Beddell, C. R., Goodford, P. J., Norrington, F. E., Wilkinson, S. and Wootton, R., "Compounds designed to fit a site of known

- structure in human hemoglobin." *Br J Pharmac* 57 201-209 (1976).
17. Ripka, W. C., Sipio, W. J. and Blaney, J. M., "Molecular modeling and drug design: Strategies in the design and synthesis of phospholipase A2 inhibitors." *Lect Heterocyc Chem* 9 S95-S104 (1987).
 18. Freudenreich, C., Samama, J.-P. and Biellmann, J.-F., "Design of inhibitors from the three-dimensional structure of alcohol dehydrogenase. Chemical synthesis and enzymatic properties." *JACS* 106 3344-3353 (1984).
 19. Varney, M. D., Marzoni, G. P., Palmer, C. L., Deal, J. G., Webber, S., Welsh, K. M., Bacquet, R. J., Bartlett, C. A., et al., "Crystal-structure-based design and synthesis of benz[cd]indole-containing inhibitors of thymidylate synthase." *J Med Chem* 35 663-676 (1992).
 20. Beddell, C. R., Goodford, P. J., Kneen, G., White, R. D., Wilkinson, S. and Wootton, R., "Substituted benzaldehydes designed to increase the oxygen affinity of human haemoglobin and inhibit the sickling of sickle erythrocytes." *Br J Pharmac* 82 397-407 (1984).
 21. Goodford, P. J., "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules." *J Med Chem* 28 849-857 (1985).
 22. Lawrence, M. C. and Davis, P. C., "CLIX: A search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure." *Proteins* 12 31-41 (1992).

23. Miranker, A. and Karplus, M., "Functionality maps of binding sites: a multiple copy simultaneous search method." *Proteins* 11 29-34 (1991).
24. Van Drie, J. H., Weininger, D. and Martin, Y. C., "ALADDIN: An integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructure searching of three-dimensional molecular structures." *J Comp-Aid Mol Des* 3 225-251 (1989).
25. Moon, J. B. and Howe, W. J., "Computer design of bioactive molecules: a method receptor-based *de novo* ligand design." *Proteins* 11 314-328 (1991).
26. DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R., "Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure." *J Med Chem* 31 722-729 (1988).
27. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. and Ferrin, T. E., "A geometric approach to macromolecule-ligand interactions." *J Mol Biol* 161 269-288 (1982).
28. Shoichet, B. K., Bodian, D. L. and Kuntz, I. D., "Molecular docking using shape descriptors." *J Comp Chem* 13 380-397 (1992).
29. DesJarlais, R. L., Sheridan, R. P., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R., "Docking flexible ligands to macromolecular receptors by molecular shape." *J Med Chem* 29 2149-2153 (1986).
30. Shoichet, B. K. and Kuntz, I. D., "Protein docking and complementarity." *J Mol Biol* 221 327-346 (1991).

31. DesJarlais, R. L., Seibel, G. L., Kuntz, I. D., Furth, P. S., Alvarez, J. C., Ortiz de Montellano, P. R., DeCamp, D. L., Babe', L. M., et al., "Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease." *PNAS* 87 6644-6648 (1990).
32. Rutenber, E., manuscript in preparation.
33. Shoichet, B. K., Molecular Docking: Theory and Application to Recognition and Inhibitor Design. University of California, 1991.
34. Gilman, A. G., Rall, T. W., Nies, A. S. and Taylor, P. (Eds.), Goodman and Gilman's The Experimental Basis of Therapeutics. (Pergamon Press, New York, 1990).

Chapter 2

Influenza Hemagglutinin: A Test Case for *De Novo* Drug Design

Influenza pandemics recur on an annual basis worldwide. Vaccination programs aimed at curtailing the spread of the disease are hampered by the fast mutation rate of antigenic sites on the virus [1]. Medical practices abroad permit use of the drugs amantadine and rimantadine to treat influenza A infections. However, due to the potential for undesirable side effects, use in the United States is recommended only for the population deemed most at risk. The cost of millions of lives and billions of dollars each winter underscores the urgent need for development of safe and effective anti-influenza drugs.

Analysis of the replication pathway of the orthomyxovirus reveals a number of steps that can be targeted for antiviral therapy. Successful infection requires host cell recognition, delivery of the infectious genome into the host cell cytoplasm, replication of the viral genes and proteins, and escape of progeny viruses. Any of these steps is potentially susceptible to intervention. However, the antiviral strategy must be specific for influenza proteins or processes in order to avoid adventitious inhibition of normal cellular functions.

Early events in the viral lifecycle leading to the deposition of the viral genes inside the cell are shown schematically in Figure 2-1.

Infection begins by binding between the hemagglutinin glycoprotein protruding from the viral envelope and sialic acid residues of cellular receptors. As the virus is endocytosed via the normal cellular pathway it encounters progressively decreasing pH. At a threshold pH specific to the particular strain of influenza, fusion between the viral membrane and the endosomal membrane is initiated [2]. This fusion event results in release of the infectious genome into the cell cytoplasm, where successive steps of the replication cycle can occur.

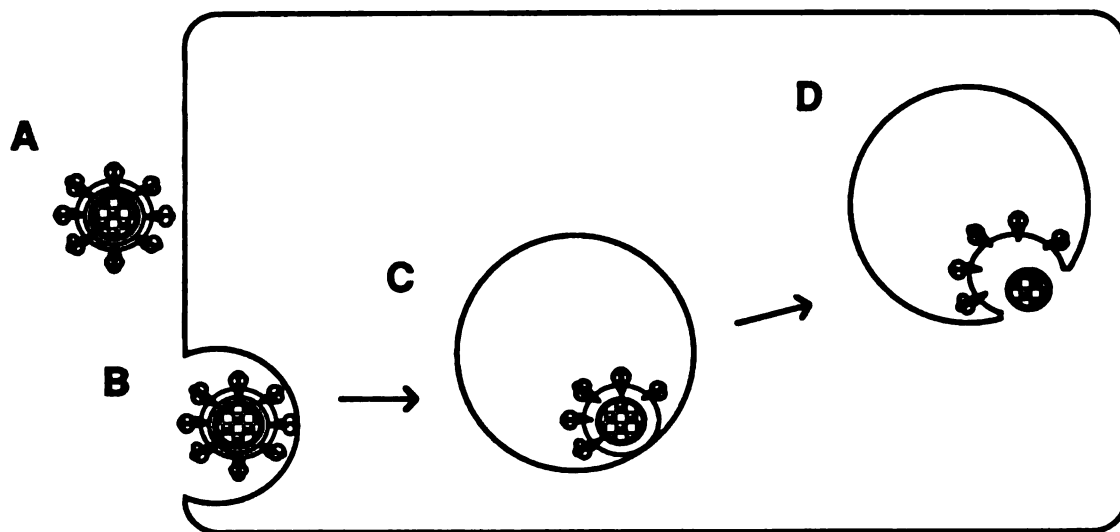


Figure 2-1. The initial steps of infection by an influenza virus. (A) The virus binds to a cellular receptor via hemagglutinin spikes projecting from the viral envelope. Binding triggers endocytosis (B). In (C) the virus has progressed along the endocytic pathway encountering successively decreasing pH environments. When a threshold pH is reached, fusion between the membrane of the virus and the membrane of the endosome occurs. Fusion results in ejection of the nucleocapsid into the cell cytoplasm (D). This figure is based on a previously published cartoon [3].

The critical role of membrane fusion in infection makes it an attractive target for inhibition. To date, this route of antiviral chemotherapy has been largely unexplored. Inhibition of fusion has the advantage of interfering with an early step in replication, prior to penetration of the virus into the host cell. As it does not aim to inhibit an enzymatic activity or to mimic any ligands, the chance of fortuitous inhibition of unintentional targets is minimized. Since fusion is a step common to the replication of all enveloped viruses, this antiviral strategy can potentially be applied to a host of other viral diseases, including those caused by togaviruses, rhabdoviruses, paramyxoviruses, herpesviruses, leukemia viruses, and retroviruses.

Examination of fusion in more detail reveals it to be a protein-mediated event triggered by the viral hemagglutinin [4]. Hemagglutinin is a trimer of identical subunits. Each monomer is composed of two polypeptide chains, HA1 and HA2, which are generated by proteolytic cleavage of a precursor, HA0. The polypeptides comprising the monomer are covalently linked by a single disulfide bond but the three monomers of a trimer are stabilized by noncovalent interactions only [5].

Residues 1-24 at the amino terminus of HA2 play a critical role in fusion. This segment, known as the fusion peptide, has been proposed to form a sided helix in which one face of the helix is composed primarily of hydrophobic amino acids [6]. While the function of the fusion peptide is not clearly understood, current evidence suggests it aids fusion by interacting with the target membrane [7, 8]. The fusion peptide is the most highly conserved region among influenza virus hemagglutinins sequenced to date

(Appendix B), and hydrophobic or sided fusion peptide sequences have been identified in the fusion proteins of a wide variety of enveloped viruses [6].

Upon exposure to low pH, hemagglutinin undergoes an irreversible conformational change that is a prerequisite for fusion. The conformational change most likely involves a rearrangement of domains rather than major secondary structural alterations since circular dichroism measurements reveal only minor differences between the neutral and low pH forms [9]. Studies of the conformational change have been facilitated by isolation of the soluble ectodomain of the integral membrane glycoprotein. This proteolytic fragment, BHA, is generated by bromelain cleavage of hemagglutinin at a site adjacent to the transmembrane domain. BHA has been identified as a reliable model for the complete protein (HA) in many assays not requiring membrane attachment [10, 11].

Previous studies on HA and BHA have shown that the low pH form of hemagglutinin¹ may be distinguished from the neutral pH form immunologically [10, 11], biochemically and biophysically. Only the low pH conformation is susceptible to cleavage by trypsin [12] and by proteinase K [11]. The low pH form of BHA has increased hydrophobic character, observed by binding to liposomes [11, 12], and partitioning into detergent solution or aggregation in aqueous solution [12]. Low pH and native hemagglutinin have also been distinguished by electron microscopy [12, 13, 14].

¹The term "hemagglutinin" is used to mean both HA, the intact integral membrane protein, and BHA, its proteolytic fragment lacking the transmembrane and cytoplasmic domains.

The crystal structure of neutral pH BHA from the X31 strain of influenza (A/Hong Kong/1968; H3N2) has been solved to 3 Å resolution [5]. Tracings of the α -carbon backbones of both the trimer and monomer are shown in Figure 2-2. Each monomer has been described as comprising three domains: a globular head domain containing the sialic acid binding site, a narrow stem composed primarily of residues of HA2, and a connecting hinge region. The fusion peptides of native hemagglutinin are buried in the trimer interface of the stem region. The conformational change is thought to release the fusion peptides from their unexposed location, freeing them to mediate fusion.

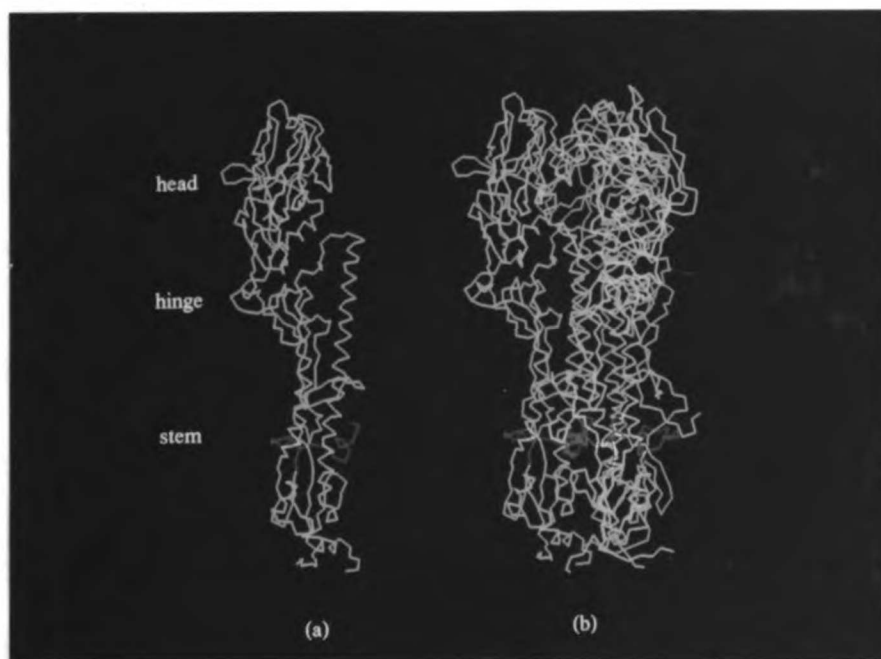


Figure 2-2. α -Carbon backbones of the BHA (a) monomer and (b) trimer from Protein Databank entry 1hmg. Fusion peptide residues 1-24 of HA2 are displayed in red. Head, hinge, and stem regions are labeled.

Since membrane fusion depends on the conformational change, inhibition of fusion peptide exposure should prevent fusion and all successive steps of viral replication. Therefore, the antiviral strategy proposed is to identify a small molecule that can bind to the native form of hemagglutinin and stabilize that conformation over any fusogenic conformation. Discovery of such an inhibitor could eventually lead to the development of a safe and effective drug.

References

1. Tannock, G. A., "Alternatives in the control of influenza." *The Medical Journal of Australia* 154 692-695 (1991).
2. Stegmann, T., Morselt, H. W. M., Scholma, J. and Wilschut, J., "Fusion of influenza virus in an intracellular acidic compartment measured by fluorescence dequenching." *BBA* 904 165-170 (1987).
3. Simons, K., Garoff, H. and Helenius, A., "How an animal virus gets into and out of its host cell." *Sci Am* 246 58-66 (1982).
4. White, J., Helenius, A. and Gething, M.-J., "Hemagglutinin of influenza virus expressed from a cloned gene promotes membrane fusion." *Nature* 300 658-659 (1982).
5. Wilson, I. A., Skehel, J. J. and Wiley, D. C., "Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution." *Nature* 289 366-373 (1981).

6. White, J. M., "Viral and cellular membrane fusion proteins." *Ann Rev Physiol* 52 675-697 (1990).
7. Stegmann, T., Delfino, J. M., Richards, F. M. and Helenius, A., "The HA2 subunit of influenza hemagglutinin inserts into the target membrane prior to fusion." *JBC* 266 18404-18410 (1991).
8. Harter, C., Bachi, T., Semenza, G. and Brunner, J., "Hydrophobic photolabeling identifies BHA2 as the subunit mediating the interaction of bromelain-solubilized influenza virus hemagglutinin with liposomes at low pH." *Biochemistry* 27 1856-1864 (1988).
9. Wharton, S. A., Ruigrok, R. W. H., Martin, S. R., Skehel, J. J., Bayley, P. M., Weis, W. and Wiley, D. C., "Conformational aspects of the acid-induced fusion mechanism of influenza virus hemagglutinin. Circular dichroism and fluorescence studies." *JBC* 263 4474-4480 (1988).
10. White, J. M. and Wilson, I. A., "Anti-peptide antibodies detect steps in a protein conformational change: low pH activation of the influenza virus hemagglutinin." *J Cell Bio* 105 2887-2896 (1987).
11. Doms, R. W., Helenius, A. and White, J., "Membrane fusion activity of the influenza virus hemagglutinin: the low pH-induced conformational change." *JBC* 260 2973-2981 (1985).
12. Skehel, J. J., Bayley, P. M., Brown, E. B., Martin, S. R., Waterfield, M. D., White, J. M., Wilson, I. A. and Wiley, D. C., "Changes in the conformation of influenza virus hemagglutinin at the pH

- optimum of virus-mediated membrane fusion." *PNAS* 79 968-972 (1982).
13. Stegmann, T., Booy, F. P. and Wilschut, J., "Effects of low pH on influenza virus. Activation and inactivation of the membrane fusion capacity of the hemagglutinin." *JBC* 262 17744-17749 (1987).
 14. Ruigrok, R. W. H., Martin, S. R., Wharton, S. A., Skehel, J. J., Bayley, P. M. and Wiley, D. C., "Conformational changes in the hemagglutinin of influenza virus which accompany heat-induced fusion of virus with liposomes." *Virology* 155 484-497 (1986).

Chapter 3

Materials and Methods

Cells, Viruses, Antibodies, and Reagents

Wt-HA-expressing CHO-DUKX cells (Wtm8005 cell line), a gift of Dr. Don Wiley, were maintained in G418 media (MEM-alpha (minimal essential media alpha) without nucleosides, 10% supplemented bovine calf serum (SCS), 2 mM glutamine, 100 U/ml penicillin, 100 mg/ml streptomycin, 600 μ g/ml geneticin (Gibco BRL), 0.3 μ M methotrexate). CV-1 cells (American Type Tissue Culture) were maintained in CV-1 growth media (DME (Dulbecco MEM), 10% SCS, 2 mM glutamine, 100 U/ml penicillin, 100 mg/ml streptomycin). MDCK2 (Madin-Darby canine kidney) cells, gifts of Dr. Barry Gumbiner, were grown in MDCK growth media (MEM-EBSS (MEM with Earle's balanced salt solution), 5% SCS, 2 mM glutamine, 100 U/ml penicillin, 100 mg/ml streptomycin, 25 mM HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid) pH 7.2). Unless otherwise noted, all tissue culture reagents were obtained from the UCSF Cell Culture Facility.

Inoculum for X31 influenza virus (H3N2) and its plaque purified subtype C22 [1] were the gift of Dr. Ari Helenius. Inoculum for influenza A/Victoria/3/75 (H3N2) was kindly provided by the Influenza Group at the Centers for Disease Control (CDC).

Drs. Richard Lerner and Ian Wilson of Scripps provided some of the α -fusion peptide antiserum. These anti-peptide antibodies were raised against residues 1-29 of HA2. The site A mouse monoclonal antibody was the gift of Dr. John Skehel. Dr. Peter Sorter provided rimantadine and Dr. Cornelia Schroeder supplied a sample of norakin hydrochloride. 5- ^3H -2'-deoxycytidine, 26 Ci/mmol, was purchased from Moravek Biochemicals (Brea, CA). ^3H -amantadine was prepared by neutron bombardment of amantadine hydrochloride (Aldrich) by Dr. C. T. Peng (UCSF) to a specific activity of 25.6 mCi/mmol.

Acquisition and dissolution of trial inhibitors

Sources for all commercial and synthesized compounds are listed in Tables 4-9, 6-11, and 7-7. Synthesis, purification, and characterization of selected compounds are summarized Appendix A. Stock solutions were prepared fresh daily. Round I compounds 1, 2, 7, 8, 14, 18, 28 were dissolved in water, 16 and 22 in 25-50% methanol/water, and 4, 5, 6, 9, 10, 11, 12, 13, 17, 20, 21, 23, 24, 30, 31, 32 in DMSO. All solutions of the Round II compounds and those numbered 83 and higher were prepared in DMSO. Compounds insoluble in DMSO or water at sufficiently high concentrations were not tested.

Preparation of viral inoculum

C22 and X31 influenza were propagated in chicken embryos as previously described [2]. Ten day old fertilized eggs were infected

with 0.01 HAU of virus in 0.1 cc sterile PBS. Infection proceeded for two days at 37°C. After incubation overnight at 4°C, the allantoic fluid was harvested sterilely then cleared of debris by centrifugation for 5 minutes at 700 x g (2000 rpm in a Beckman Accuspin).

Virus purification

Virus was purified from inoculum as previously described [2]. Fertilized eggs were infected as above. Following a 30 minute debris spin, the virus was pelleted at 40,000 x g (Type 19 rotor at 17K rpm) for 2 hours 40 minutes. Pellets were eluted overnight in PBS then dounced. The slurry was incubated at 37°C for 15 - 30 minutes, then spun for 5 minutes at 1000 x g (2500 rpm in a Beckman Accuspin) to pellet red blood cells and aggregates. The virus was layered over a 30%/60% w/vol sucrose/PBS step gradient and spun at 75,000 x g (SW27, 24K rpm) for 90 minutes. The interface band was collected and pelleted in PBS under the same conditions. Pellets were eluted overnight at 4°C then dounced. Aggregates were cleared as above. Yield of viral protein was determined by Lowry assay.

Purification of unlabeled BHA

Bromelain-digested hemagglutinin trimers were isolated from influenza virions following a published procedure [3]. Approximately 10 mgs purified virus were diluted to 5 mg/ml in 0.1 M Tris HCl, pH 8.0 then incubated overnight at 37°C with 1.25 mg/ml bromelain (Sigma) and 50 mM β -mercaptoethanol (Biorad). Bromelain was

inactivated by the addition of 1 mM N-ethylmaleimide (Sigma). The viral cores were pelleted at 4°C in 0.1 M Tris at 100,000 x g (SW41, 25K rpm) for 1 hour. The supernatant was concentrated then loaded on a 5%-25% w/vol sucrose/PBS gradient and centrifuged at 4°C for 16 hr at 160,000 x g (SW41, 37K rpm). Approximately 25 fractions were collected, and protein was detected by the Biorad procedure. Peak fractions corresponding to 9S trimers were pooled.

BHA iodination

Protocol 1: IODOGEN

One iodogen-coated (Pierce) tube prepared following the manufacturer's directions was used to mix 50 µg BHA with 0.5 mCi Na¹²⁵I (Amersham) in PBS in a total of 100 µl. The mixture was incubated for 5 minutes at RT then applied to a washed Dowex (Sigma) column. The column was rinsed three times with PBS. Specific activity of each fraction collected was determined in a gamma counter. BHA trimers were isolated by sucrose gradient centrifugation as described above.

Protocol 2: IODOBEADS

One washed iodobead (Pierce) was incubated with 100 µl PBS and 0.5 mCi Na¹²⁵I (Amersham) for 5 minutes at RT. 50 µg BHA was added and the reaction allowed to proceed for 5 minutes at RT.

Iodinated trimers were isolated by Dowex column chromatography and sucrose gradient centrifugation as in Protocol 1.

Purification of ^{35}S -Met-HA

Metabolically labeled ^{35}S -Met-HA was prepared similarly to a published protocol [4]. T-150 flasks of 60-70% confluent CV-1 cells were infected with 6 ml of 10,000 HAU/ml inoculum in DME without serum. Cells were incubated for 90 minutes at 37°C, 5% CO₂ with periodic agitation. After addition of CV-1 growth media, incubation continued for a further 4 hr. The cells were then washed with Met-media (MEM (GIBCO), 2.2 g/L NaHCO₃, 58 mg/l lysine, 53 mg/l leucine, pH 7.4) and incubated for 90 min at 37°C, 5% CO₂. Labeling continued overnight in Met-media with 2.5% SCS and 1 mCi/flask ^{35}S -methionine (ICN). Cells were washed and collected by scraping in PBS, pelleted for 5 minutes at RT at 200 x g, then lysed on ice for 15 minutes in 3.5 mls cell lysis buffer (10 mM MES (2-[N-Morpholino]ethanesulfonic acid), 10 mM HEPES, 100 mM NaCl, 0.1% NP40 (nonidet-P40), pH 7.1) containing 1 mM PMSF (phenyl methyl sulfonyl fluoride) and 10 µg/ml aprotinin. Cell debris was pelleted by centrifugation for 30 minutes at 70,000 x g (TLA 100, 40K rpm). The HA precursor, HA0, was activated by digestion for 10 minutes at RT with 10 µg/ml N-tosyl-phenylalanine chloromethylketone (TPCK) trypsin (Sigma). After addition of 10 µg/ml soybean trypsin inhibitor (STI) (Sigma), the protein was applied to a ricin-agarose column. Following extensive washing at 4°C with PBS/0.1% NP40 the HA was eluted with 0.2 M galactose/PBS/0.1% NP40. Peak fractions

were pooled, concentrated, and purified on a 5%-25% w/vol sucrose/MES Saline (0.13 M NaCl, 20 mM MES, pH 7.0)/0.1% NP40 gradient for 16 hours at 170,000 x g (SW41, 37K rpm) and 4°C. Fractions corresponding to the 9S trimer peak were pooled.

Purification of ^3H -Leu-BHA

Metabolically labeled ^3H -Leu-BHA was prepared similarly to ^{35}S -Met-HA. Leu- media (MEM (GIBCO), 2.2 g/l NaHCO_3 , 58 mg/l lysine, 15 mg/l methionine, pH 7.3) was used in place of Met- media and infected cells were labeled with 0.5 mCi/flask 3,4,5- ^3H -leucine (NEN). The trypsin digestion was replaced by cleavage with 0.1 mg/ml bromelain, 20 mM β -mercaptoethanol for 16 hours at RT. The reaction was quenched by addition of 1 mM N-ethylmaleimide and BHA was purified by ricin affinity chromatography and sucrose gradient purification as above.

Scintillation proximity assay (SPA)

^3H -Leu BHA was diluted to 10,000 cpm (counts per minute) per 200 μl in MSSH buffer (10 mM HEPES, 10 mM MES, 10 mM succinate, 0.10 M NaCl, pH 7.0) containing 0.1% NP40. Protein was incubated with the specified concentration of trial compound for 25 minutes at RT. The same volume of solvent used to dissolve the compound was added to control samples. For the Round II compounds and the derivatives of 83, the final concentration of DMSO in all solutions was 0.67% v/v. A predetermined amount of 1N HAc was added to bring

the reactions to the appropriate pH (usually pH 5.0). Following a 5 minute (unless specified otherwise) incubation at RT, reactions were reneutralized with 1N NaOH. For POST controls, compounds were added to the appropriate concentration after reneutralization. 200 μ l aliquots of the protein solution were added to scintillation vials containing 100 μ l protein A-SPA beads (Amersham), 5% fetal bovine serum (FBS), and antibody. Vials were incubated overnight at RT with constant shaking. Each reaction used 10 μ l of fusion peptide antiserum diluted 1:10 with 50% glycerol. Cpm were detected in a Beckman LS3801 scintillation counter without addition of liquid scintillant.

Dilution experiment

3 H-Leu BHA at 1000 cpm/ μ l in MES Saline/0.1% NP40 was incubated with 1 mM compound 83 in DMSO or DMSO alone. The concentration of DMSO was 2.2% v/v. After incubation for 2 hours at RT, the samples were diluted 25x with MES Saline/0.1% NP40 containing either 1 mM 83/DMSO or DMSO alone. Samples were incubated a further 2 hours, then acidified to pH 5.0 for 5 minutes at RT and reneutralized. POST controls were then mixed with the appropriate concentration of compound. The equivalent amount of DMSO was added to all other samples. Reactions were precipitated as described above.

Calculation of nonspecific quenching

Apparent inhibition due to nonspecific effects was quantitated from the POST controls with the equation:

$$\text{observed cpm} = \text{true cpm} * \text{quench factor}$$

where the quench factor is the fraction by which the compound reduces the cpm of the POST control compared to otherwise identical samples with compound omitted. The true cpm is the cpm that would be observed if quenching did not occur.

Calculation of % inhibition

Inhibition was calculated as the % difference in cpm precipitated by a given sample compared to the cpm precipitated by the appropriate control sample. The % inhibition compensated for quenching was calculated as:

$$\text{true inhibition} = \left(1 - \frac{(1 - \text{inhibition})}{(1 - \text{inhibition}_{\text{POST}})} \right) \times 100\%$$

where inhibition is the observed inhibition in compound-containing samples and inhibition_{POST} is the observed inhibition of compound-containing POST samples.

Calculation of IC₅₀

The IC₅₀, the concentration of ligand producing 50% specific inhibition, was determined for each of the derivatives of compound 83. The maximum inhibition corrected for quenching was $63.4 \pm 10.0\%$. Therefore, half-maximal inhibition occurs at 31.7%. The concentration at which each compound reduces the cpm precipitated by approximately 31.7% was determined from a plot of log concentration vs. % inhibition.

pH control

This control tests whether the trial compounds altered the pH of the solutions. The acidification reactions were identical to the acidification reactions performed for SPA analysis, except that the detergent and radiolabeled protein were omitted. The pH of acidified compound-containing solutions was compared to that of the analogous DMSO-containing solution.

Immunoprecipitation

Protocol 1

¹²⁵I-BHA was diluted with MES Saline/0.1% NP40 to 10,000 cpm/100 μ l. Compounds were added to the appropriate final concentration and the mixture incubated for 15 minutes at RT.

Following acidification to pH 5.0 (unless otherwise indicated) with 1N HAc, the reaction was incubated for 5 minutes at RT, then reneutralized by addition of 1N NaOH. For reactions at pH 7, MES Saline was added in place of HAc and NaOH.

100 μ l aliquots of each reaction were incubated overnight at 4°C or at RT for 2 hours with the α -fusion peptide antibody diluted to 1:40. Antibodies were bound to 25 μ l of 10% zysorbin for 45 minutes at 4°C with constant shaking. Complexes were washed twice with BHA wash buffer (0.5 M NaCl, 0.1 M Tris, pH 8.0) then precipitated by a 5 minute spin at 10,000 x g. Pellets were counted in a gamma counter.

Protocol 2

Radiolabeled HA or BHA was diluted with MES Saline/0.1% NP40 to 10,000 cpm/100 μ l. Compounds in DMSO (or DMSO alone) were added to the appropriate final concentration and incubated for 20 minutes at RT. 1N HAc was added to bring the protein solution to the appropriate pH (pH 5.0 unless otherwise indicated). After incubation at RT for 5 minutes (unless specified), the reactions were reneutralized by addition of 1N NaOH. For neutral pH reactions, the same volume of MES Saline was added in place of HAc and NaOH.

Samples were split into 100 μ l aliquots. The α -fusion peptide antibody was added to a final dilution of 1:5 and incubated at RT for 2 hours. 75 μ l of 25% protein A-agarose were then added and the mixture incubated with constant shaking for 1 hour at RT. Beads were pelleted by a 5 minute spin at 10,000 x g and then washed

twice with BHA wash buffer. Pellets were resuspended in 50 μ l 10% sodium dodecyl sulfate, and the tubes washed twice with 100 μ l scintillation fluid. An additional 3 mls of scintillation fluid prior to scintillation counting.

Proteinase K digestion

Protocol 1: I-BHA

125 I-BHA was mixed with MES Saline and 0.5% NP40 at a concentration of 10,000 cpm/100 μ l and incubated with the appropriate compounds for 10 minutes at RT. The pH was dropped to 5.0 by the addition of 1N HAc. After incubation for 5 minutes at RT the reaction was reneutralized with 1N NaOH. For POST controls, the appropriate concentration of compound was added after reneutralization. The protein solution was digested in PBS with 0.25 mg/ml proteinase K (Sigma) for 30 minutes at 37°C. The reaction was quenched by transferring the tubes to 0°. Undigested protein was precipitated on ice for 2 hours in the presence of 0.025 μ g/ml salmon sperm DNA and 10% TCA (trichloroacetic acid). Pellets were washed twice with 10% TCA then counted in a gamma counter.

Protocol 2: S-HA

35 S-HA was diluted with MSSH Buffer and 0.5% Triton X-100 (or NP40) to 10,000 cpm/100 μ l and incubated with the appropriate compounds for 25 minutes at RT or overnight at 4°C. The pH was

dropped to 5.0 by the addition of 0.5 M succinate. After 15 minutes at RT reactions were reneutralized with 1N NaOH. For POST controls, the appropriate concentration of compound was added after reneutralization. Digestion, precipitation, and counting were done as in Protocol 1.

Calculation of % inhibition

The cpm precipitated by duplicate samples were averaged. The % inhibition for each compound was determined by comparing the number of cpm precipitated in the presence of the ligand to the cpm precipitated in control samples with test ligands omitted. The cpm observed in control samples maintained at pH 7 represents the maximum number of cpm precipitable, or 100% inhibition. The number of counts precipitated by controls at pH 5 was taken as 0% inhibition (maximum digestion). % Inhibition for the compound-containing samples was calculated by linear extrapolation between the 0 and 100% values. For compounds tested in multiple independent experiments, the % inhibition from each experiment was averaged.

Aggregation assay

¹²⁵I-BHA was diluted with MES Saline to 10,000 cpm/100 μ l. Test compounds were added and incubated for 15 minutes at RT. The mixture was acidified to pH 5.0 by addition of a predetermined amount of 1N HAc, then incubated at RT for 5 minutes. After

reneutralization, BHA trimers are separated from the heavier BHA aggregates on 5%-20% w/vol sucrose/PBS gradients at 4° and 85,000 x g (SW55, 30K rpm) for 16 hours. Fractions were collected and processed for scintillation counting. POST controls, in which the trial ligands were added following the acidification reaction, tested whether the compounds dissassociated aggregated trimers of pH 5 BHA.

To compute % inhibition, the background cpm determined from counting scintillation fluid alone were subtracted from the cpm of each fraction. The counts in the fractions representing the unaggregated trimer peak were summed. The total cpm of the pH 7 control sample without test ligands represented 100% inhibition, and that of the control sample at pH 5 represented 0% inhibition. The % inhibition of all other samples was calculated by linear extrapolation.

Hemagglutination

Serial 2-fold dilutions of inoculum in PBS were placed in a 96 well plate with V-shaped wells. An equal volume of 0.5% washed human red blood cells in PBS was added to each. Wells were covered and incubated at 4°C. After 4 hours the last well agglutinated was recorded.

Hemagglutinating units (HAU) are computed by the formula:

$$\text{HAU/ml} = 20 \times 2^{(n-1)}$$

where n is the last well agglutinated.

In testing compounds for inhibition of hemagglutination, the virus was incubated for 30 minutes at RT with 2x concentrations of ligand. Serial 2x dilutions of virus were done in a 2x solution of compound/PBS. Addition of the rbc's halved the compound concentration in each well.

Hemolysis

Protocol 1

Human red blood cells (rbc's) were washed twice with PBS then suspended to 1% v/v with MES Saline. Approximately 1.5 μ g X31 virus was mixed with trial compound and MES Saline in a final volume of 25 μ l then incubated for 25 minutes at RT. Following addition of 225 μ l 1% rbc's, the reaction was warmed to 37°C for 5 minutes. The pH was lowered to 5.0 (unless otherwise noted) by addition of a predetermined amount of 1N HAc. The reaction was reneutralized with 1N NaOH after incubation for 5 minutes at 37°C. Intact rbc's were pelleted by centrifugation at 10,000 x g for 3 minutes. The OD₅₇₀ was measured on 100 or 150 μ l aliquots of supernatant. Background values were determined from identically treated samples lacking virus. Titrations demonstrated that the amount of virus used fell in the middle of the linear range.

Protocol 2

Protocol 2 is similar to Protocol 1 except that virus was preincubated with test compounds for 45 minutes at RT and rbc's were diluted from an 8% suspension. Reactions were incubated at low pH for 15 minutes at 37°C.

Protocol 3

Protocol 3 is similar to Protocol 1 except that the C22 strain of influenza X31 was used. Virus titrations determined that 0.6 µg of viral protein (by Lowry) gave an OD in the center of the linear range. Virus and compound were preincubated for 0.5 hr at RT in a volume of 100 µl. 350 µl of 1% washed rbc's were added. Warming, acidification, and reneutralization were done as in Protocol 1 except that the reaction was allowed to proceed for 15 mins at 37°C. OD₅₇₀ was determined as above.

Hb controls were identical to other reactions except that virus was replaced by the same volume of assay buffer and pre-lysed rbc's were used instead of intact rbc's. Rbc's were lysed in 10 mM MES, then centrifuged at 10,000 x g for 5 minutes. The supernatant was diluted 1:20 with MES Saline.

Calculation of IC₅₀

The % reduction on OD in sample containing compounds were calculated as:

$$\left(1 - \frac{(OD_{DMSO,-} - OD_{DMSO,+})}{(OD_{-} - OD_{+})} \right) \times 100\%$$

where $OD_{DMSO,+}$ is the absorbance of samples containing DMSO (or the appropriate solvent), $OD_{DMSO,-}$ the background lysis of rbc's without virus present but with DMSO, OD_{+} is the absorbance of the sample and virus containing samples and OD_{-} the background lysis caused by the compound in the absence of virus. The IC_{50} is defined as the concentration reducing the OD by the same amount as samples containing half the amount of virus. Depending on the day, 50% less virus reduced the OD by 25-60%.

Fluorescence dequenching fusion assay

Labeling rbc's with octadecyl rhodamine

Octadecyl rhodamine (R_{18}) was incorporated into the membrane of rbc's as described previously [5]. 30 μ l of 1 mg/ml R_{18} in ethanol was added to 20 ml of 1% washed human rbc's in RPMI (Sigma) while vortexing. The mixture was incubated in the dark for 15 minutes at RT. 30 mls RPMI/10% SCS was added and the mixture incubated in the dark for a further 20 minutes. The cells were washed 6 times, then resuspended to 10 mls with RPMI. The emission (F) at 590 nm (excitation at 560 nm) of 50 μ l cells in 3 mls RPMI was measured, then 50 μ l 10% NP40 was added and the

emission (F_{NP40}) measured following an overnight incubation. The fluorescence dequenching is defined as:

$$\%FDQ = \frac{(F_{NP40}-F)}{F_{NP40}} \times 100\%$$

Values of 75% - 80% were typically observed.

Binding labeled rbc's to HA-expressing cells

90 mm dishes of wtm-8005 cells at 80-90% confluency were washed with RPMI. Cell surface HA0 was cleaved with RPMI, 10 μ g/ml TPCK trypsin, 0.2 mg/ml neuraminidase (Sigma) for 10 minutes at RT. For HA0 controls, 10 μ g/ml STI was substituted for trypsin. The plates were washed with RPMI, 10% SCS, 20 μ g/ml STI, then with RPMI alone. 5 ml 0.1% R₁₈-rbc's were added and incubated for 10 minutes at RT with occasional swirling. The unbound rbc's were removed with 6 RPMI washes. Cells were removed from the dishes with PBS, 0.5 mM EDTA, 0.5 mM EGTA and pelleted in MSSH buffer. Cells were resuspended to 0.5 ml/dish with MSSH buffer.

Fusion

HA-expressing wtm cells with prebound R₁₈-rbc's were equilibrated to 37°C in fusion buffer. Data was collected for 2 minutes both before addition of test compound and after incubating the cells with the compound for 5 minutes. At time t=0, the pH was

lowered to 5.2 by the addition of a predetermined volume of 1M citric acid. Data was acquired for a further 5 minutes. The fluorescence dequenching at infinite dilution was determined following an overnight incubation in 0.5% NP40. All fluorescence experiments were conducted at 37°C with constant stirring in an LS-5B fluorimeter.

The fluorescence dequenching was determined from the equation:

$$\%FDQ = \frac{(F_t - F_0)}{(F_{NP40} - F_0)} \times 100\%$$

where F_t is the fluorescence observed at time t , F_0 is the fluorescence detected in the presence of the compound immediately before addition of acid, and F_{NP40} is the fluorescence of the NP40 solution.

Syncytia formation: fusion from without

Influenza-mediated cell-cell fusion was induced following the previously described protocol for fusion from without [1]. 50-60% confluent plates of CV-1 cells were washed with cold PBS. A total of 1 ml DME, 100 mg/ml streptomycin, 100 U/ml penicillin (DME/P&S) with or without 6.5 μ g of virus (by Lowry protein determination) and the appropriate concentration of trial compound was added to each well on ice. The final concentration of DMSO in all wells was 0.1% by volume. The plates were incubated at 4°C for 1 hour with gentle agitation. Following the binding period, the virus solution was

aspirated and replaced with prewarmed MSSH buffer, pH 5.2, containing compound and DMSO. The buffer was aspirated after incubation at 37°C for 3 minutes. Cells were allowed to recuperate in CV-1 growth media for 4 hrs at 37°, 5% CO₂. Cells were examined for any alterations in cell morphology then stained for 30 minutes at RT with 0.2% crystal violet in 50% ethanol.

EIA antiviral assay

This protocol is adapted from the CDC protocol. MDCK2 cells were seeded at 35,000 cells per well in 96 well cluster dishes and grown in MDCK growth media for 24 hrs at 37°C, 5% CO₂. C22 virus was preincubated with the appropriate concentration of test compound in MEM-EBSS, 100 mg/ml streptomycin, 100 U/ml penicillin for 25 mins at RT. All samples contained 0.67% v/v DMSO and 0-1.5 HAU virus per 100 µl. Cells were infected with 100 µl virus/compound/MEM-EBSS for 15 hrs at 37°C, 5% CO₂. Trypsin, normally included to cleave HA0 to HA, was omitted so only a single round of infection could occur. Cells were rinsed once with PBS then fixed with 80% acetone/PBS for 15 min at RT and allowed to air dry. Plates were washed with PBS/0.05% Tween 20 then blocked with EIA diluent (PBS, 1% FBS, 0.1% Tween 20) for 30 min at RT. Wells were incubated for 1 hr at 37°C with site A monoclonal antibody diluted to 1:2000 with EIA diluent, washed 4 times with PBS/0.05% Tween 20, then incubated for 1 hr at 37°C with F(ab)'-2-goat-anti-mouse Ig(G)-peroxidase (Boehringer Mannheim) diluted to 1:6000 with EIA diluent. Plates were washed 4 times more with PBS/0.05%

Tween 20. A 0.3 mg/ml solution of 3,3',5,5'-tetramethylbenzidine substrate (Sigma) was prepared in citrate-acetate buffer (0.1 M sodium acetate brought to pH 5.5 with 1.0 M citric acid) with 6% v/v DMSO and .005% H₂O₂. Color development was allowed to proceed for 10 min at RT, then was stopped by the addition of 2 M sulfuric acid. The OD₄₁₀ was measured for each well.

All samples were done in triplicate, and each concentration of trial compound was tested with at least two different concentrations of virus. A standard curve relating the amount of infecting virus (preincubated with the appropriate solvent) to OD₄₁₀ was determined from virus titrations on each plate. The amount of virus producing an OD equivalent to that observed in each compound-containing well was determined from the standard curve. % Inhibition is expressed as the % difference between that amount of virus and the input amount. The % inhibition was plotted as a function of log concentration. The IC₅₀, the concentration at which infectivity is reduced by 50%, was read off the graph.

MTT viability assay

Protocol 1

Infections of MDCK2 cells with virus, compounds, and DMSO were set up identically as for the EIA assay. Following the 15 hour incubation, test compounds were washed away with PBS. 50 µl of MDCK2 growth media and 25 µl 2 mg/ml 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (Sigma) in PBS were added to

each well. Plates were returned to 37°C, 5% CO₂ for 4 hrs. Crystals were broken up by vigorous mixing following addition of 80 µl isopropanol/.04 N HCl. After incubation for 30 mins at RT, OD₅₇₀ was read on an ELISA reader. % Inhibition is the % reduction of OD₅₇₀ in the wells infected with compound compared to otherwise identical samples with compound omitted. This assay has been published elsewhere [6].

Protocol 2

CV-1 cells were plated in 96 well tissue culture dishes at 10,000 cells per well and grown in CV-1 growth media for 24 hrs. The appropriate concentration of compound in DME/P&S with 0.6% DMSO was added and incubated for 1 hour at 4°C with constant tilting. Following the binding period, cells were examined for any observable changes in morphology. Cells were rinsed twice with PBS then incubated with 25 µl 2 mg/ml MTT in PBS and 50 µl DME/P&S. Plates were incubated and processed for spectrophotometry as in Protocol 1.

Equilibrium dialysis

Precut dialysis membranes (Hoeffer) with a 12,000-14,000 MW cutoff were prepared for dialysis following the manufacturer's directions. Membranes were rinsed with methanol, then boiled for 5 minutes in 200 mls of ddH₂O containing 10 g Na₂CO₃ and 3.72 g di Na EDTA. The membranes were then washed three times with distilled

water, and stored in assay buffer (MES Saline) for at least 48 hours at 4°C.

The equilibrium dialyzer apparatus (Hoeffer) was used with teflon disks containing 8 pairs of 50 μ l chambers. 50 μ l of radiolabeled ligand in assay buffer were loaded on one half-well, and 50 μ l of radiolabeled ligand, cold BHA, and assay buffer in the other half-well. A 1 mm glass bead presoaked in assay buffer was placed in each chamber to mix the solutions. The injection ports were sealed, then the wells were rotated for 20 hours at RT, conditions empirically found to be sufficient to reach equilibrium. Four 10 μ l aliquots were removed from each half-well and processed for scintillation counting. Results from identical reactions were averaged. Binding is expressed as the difference in cpm between the halves containing protein and ligand and the halves containing ligand alone.

Gel filtration chromatography

An sephadex G-25 1.7 ml column (Isolab) was pre-equilibrated with 3-4 column volumes of assay buffer (MES Saline). The appropriate concentration of radiolabeled ligand was incubated with assay buffer with or without unlabeled BHA for 1 hour at RT, then applied to the top of the column. Fractions were collected and counted in 3 mls ECOLITE in scintillation fluid.

DOCK

Overview

The DOCK package of computer programs identifies invaginations on a receptor surface, orients small molecules in a site of interest, and scores each ligand by the degree it complements the shape and/or electrostatic properties of the site. DOCK requires three-dimensional coordinates of both ligand and receptor, and treats both molecules as rigid bodies. Summaries of the programs comprising the DOCK package and the differences between versions 1, 2 and 3 are presented here. The algorithms are described in detail elsewhere [7, 8, 9, 10].

Site Characterization: Sphere Generation

The program SPHGEN [7] builds spheres analytically from the molecular surface of a receptor. The radii of the spheres is proportional to the concavity of the surface: flat regions are represented by larger spheres while small spheres are generated in highly featured regions. A set of heuristics reduces the number of spheres from one per surface point to one per site atom. A single linkage algorithm identifies sets of overlapping spheres ("clusters"), each of which represents a potential binding site.

In DOCK2.0 and later versions, the option to tailor the clusters has been included. The program CLUSTER [9] allows the user to

define alternate means of paring the unpruned set of SPGHEN spheres.

Generation and Evaluation of Ligand Orientations

DOCK [7, 8, 9] uses the sphere definition of a target site to guide the positioning of ligands. While the details of the algorithm have changed significantly between versions 1 and 2 the basic principle has been maintained. Sphere centers indicate possible positions for ligand atoms in the site. Selected ligand atoms are mapped onto subsets of sphere centers with internal distances approximating those of the ligand atoms. Each mapping of ligand atoms onto appropriate sphere centers defines a transformation matrix that orients the small molecule in the site. A minimum of four atom-sphere center pairs are required to define a unique configuration. Although thousands of orientations are typically found for each small molecule, the search is non-exhaustive.

Following transformation of the ligand coordinates into the receptor site, the orientation is scored by evaluating the extent of shape complementarity between the receptor and the ligand. The scoring function approximates a van der Waals interaction energy. Ligand atoms docked within attractive distances of receptor atoms are assigned positive scores. Orientations in which ligand atoms overlap receptor atoms are assigned negative scores and discarded. In DOCK1, the score for each orientation is a function of the pairwise distances between each ligand atom and each receptor atom. In DOCK2 and later versions, the program DISTMAP pre-scores a grid

over the receptor site [9]. The score for each orientation is the sum of the scores of the grid points nearest each ligand atom. DISTMAP allows different distance parameters to be used for polar and nonpolar site atoms.

DOCK2.1 and later versions evaluate electrostatic complementarity in addition to shape complementarity. In DOCK2.1, the electrostatic potential at each grid point is calculated using DELPHI [11, 12]. Each small molecule is assigned both a shape score and an electrostatic score. DOCK3.0 [10] uses an AMBER-derived force field [13] to evaluate both electrostatic and van der Waals interactions.

DOCK may be run in two modes, SEARCH and SINGLE. In SINGLE mode one ligand is oriented in the site and every acceptable configuration is retained. In SEARCH mode, DOCK reads a database of structures. Each ligand is docked in thousands of orientations, but only the highest scoring orientation is saved. The ligands are ranked by their top score, and a list of the highest scoring ligands is produced. In DOCK2.1, the orientations receiving the top shape score and those receiving the best electrostatic scores are saved independently.

Databases

The database new.db [14] contains 10,000 compounds representative of the variety of molecular shapes in the Cambridge Crystallographic Database (also known as the Cambridge Structural Database) [15]. The coordinates are all experimentally determined.

MACCS (Molecular Design, Ltd.) provides CONCORD (Tripos)-generated structures of compounds from the Fine Chemicals Directory, Comprehensive Medicinal Chemistry, and Molecular Drug and Data Report. Revision 1.0 and the 1989 databases were used. Charges were calculated [10] for each of the chemicals using the method of Gasteiger and Marsili [16] as incorporated into SYBYL (Tripos).

Hardware and other Software

Molecular surfaces were computed on SUN3 workstations with DMS [17], a distributed processing implementation of Connolly's molecular surface algorithm [18]. Energy minimization was done with AMBER 3.0, revision 3A [19]. The 1989 Aldrich catalog was searched online with ALDRICHEM Data Search (Aldrich). Computer graphics software MIDAS [20] and its successor, MIDASPLUS, were run on Silicon Graphics IRIS workstations.

References

1. Doms, R. W., Gething, M.-J., Henneberry, J., White, J. and Helenius, A., "Variant influenza virus hemagglutinin that induces fusion at elevated pH." *J Virol* 57 603-613 (1986).
2. Skehel, J. J. and Schild, G. C., "The polypeptide composition of influenza A viruses." *Virology* 44 396-408 (1971).

3. Doms, R. W., Helenius, A. and White, J., "Membrane fusion activity of the influenza virus hemagglutinin: the low pH-induced conformational change." *JBC* 260 2973-2981 (1985).
4. White, J. M. and Wilson, I. A., "Anti-peptide antibodies detect steps in a protein conformational change: low pH activation of the influenza virus hemagglutinin." *JCB* 105 2887-2896 (1987).
5. Morris, S. J., Sarkar, D. P., White, J. M. and Blumenthal, R., "Kinetics of pH-dependent fusion between 3T3 fibroblasts expressing influenza hemagglutinin and red blood cells. Measurement by dequenching of fluorescence." *JBC* 264 3972-3978 (1989).
6. Mosmann, T., "Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays." *J Immunol Methods* 65 55-63 (1983).
7. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. and Ferrin, T. E., "A geometric approach to macromolecule-ligand interactions." *J Mol Biol* 161 269-288 (1982).
8. DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R., "Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure." *J Med Chem* 31 722-729 (1988).
9. Shoichet, B. K., Bodian, D. L. and Kuntz, I. D., "Molecular docking using shape descriptors." *J Comp Chem* 13 380-397 (1992).
10. Meng, E. C., Shoichet, B. K. and Kuntz, I. D., "Automated docking with grid-based energy evaluation." *J Comp Chem* 13 505-524 (1992).

11. Gilson, M. K. and Honig, B. H., "Energetics of charge-charge interactions in proteins." *Proteins* 3 32-52 (1988).
12. Gilson, M. K., Sharp, K. A. and Honig, B. H., "Calculating the electrostatic potential of molecules in solution: method and error assessment." *J Comp Chem* 9 327-335 (1987).
13. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr and Weiner, P., "A new force field for molecular mechanical simulation of nucleic acids and proteins." *JACS* 106 765-784 (1984).
14. Seibel, G. L., personal communication.
15. Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., et al., "The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information." *Acta cryst B* 35 2331-2339 (1979).
16. Gasteiger, J. and Marsili, M., "Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges." *Tetrahedron* 36 3219-3288 (1980).
17. Huang, C., personal communication.
18. Connolly, M. L., "Solvent-accessible surfaces of proteins and nucleic acids." *Science* 221 709-713 (1983).
19. Seibel, G. L., Singh, U. C., Weiner, P. K., Caldwell, J. and Kollman, P. A., University of California, San Francisco, 1990.
20. Ferrin, T. E., Huang, C. C., Jarvis, L. E. and Langridge, R., "The Midas display system." *J Mol Graph* 6 13-27 (1988).

Chapter 4

Site Selection and Inhibitor Design Round I

Development of an inhibitor of the conformational change of influenza hemagglutinin (HA) by a structure-based method requires both computer-aided selection and experimental testing of candidate ligands. This chapter describes the discovery of potential binding sites for such inhibitors and the first round of ligand design, in which DOCK was used to identify molecular frameworks with shape complementarity to the target site. Searching databases of chemical compounds in Round II of the project is detailed in Chapter 6. The experimental results from biochemically screening the potential inhibitors selected in Rounds I and II are discussed in Chapters 5 and 7, respectively.

Computational identification of potential binding sites requires calculation of the molecular surface of the receptor from its three-dimensional structure. Crystallographic coordinates of the soluble ectodomain of X31 hemagglutinin (BHA) have been deposited as entry 1hmg [1] in the Brookhaven Protein Databank (PDB) [2, 3]. The structure includes all six polypeptide chains in the trimer. HA1 and HA2 of the first monomer are labeled with chain identifiers A and B, respectively. Chains C, D and E, F represent HA1 and HA2 of the two other monomers in the trimer. Even with non-polypeptide residues such as water molecules and carbohydrate moieties omitted, the

1hmg structure has 11871 non-hydrogen atoms, too many for a reasonable molecular surface calculation. To overcome the time constraints and disk usage limitations, BHA was split into overlapping thirds. The residues of the A/B monomer assigned to each section are listed in Table 4-1. Initially just the surface contributed by this monomer was examined; the interfacial surfaces were tested later (see below).

section	residues	number of surface points
"head"	55A-275A	44303
"hinge"	20A-68A,80A-93A,103A-116A,260A-321A,49B-111B	45527
"stem"	1A-39A,315A-328A,1B-50B,100B-175B	39540

Table 4-1. Definition of sections for the initial molecular surface calculation. Surface points were computed with the program DMS at a density of 5.4 points/Å² with a probe radius of 1.4 Å. Water molecules and carbohydrate moieties were omitted from the calculation. The surface contributed by residues in each section was computed in the context of the entire trimer.

The program SPHGEN [4] was used to identify potential binding sites from the molecular surface of each section. SPHGEN describes each invagination as a set ("cluster") of overlapping spheres. Each cluster represents the negative image of a potential binding site. The

sphere representation will be used by DOCK to guide the positioning of ligands in the site. Parameter values used in the sphere generation are listed in Table 4-2. The SPHGEN algorithm has been described previously and is summarized in Chapter 3.

variable	variable name	value
density type	<i>dentag</i>	X
surface normal dot product limit	<i>dotlim</i>	0.0
maximum sphere size	<i>radmax</i>	5.0

Table 4-2. Values of parameters used in sphere generation. Spheres were calculated with the program SPHGEN from the DOCK1.1 release.

Table 4-3 summarizes the results of the sphere calculation. The head region yielded 10 clusters composed of at least 15 spheres, the hinge region gave 12, and the stem region 6. Smaller clusters were deemed uninteresting for inhibitor design. Cluster 1 of the head region includes the sialic acid binding site.

Biological, physical, and chemical properties were considered in selecting the most interesting site from all the clusters. The target site must satisfy the goal of inhibiting fusion peptide exposure. In theory, a compound bound to a site directly abutting on the fusion peptide could lock that segment in a nonfusogenic conformation. In contrast, it was unknown whether binding to a site distant from the fusion peptide region would be able to inhibit the changes in the

stem¹. Therefore, only sites in the stem region were initially considered, since that region includes residues of the fusion peptide and thus seemed most likely to contain a site that could elicit the desired biological activity.

cluster number	number of spheres		
	"head"	"hinge"	"stem"
1	110	71	153
2	50	50	62
3	47	45	30
4	40	45	27
5	36	36	17
6	26	35	15
7	18	30	
8	17	19	
9	16	19	
10	16	15	
11		15	
12		15	

Table 4-3. SPHGEN clusters computed from the molecular surfaces of the three sections comprising the A/B monomer (Table 4-1). A total of 28 clusters were found for the "head" region, 27 for the "hinge", and 23 for the "stem". Run times were 5:43 and 7:05 hours of cpu time on a VAX 8650 for the stem and head regions, respectively. Clusters are numbered consecutively by the number of spheres, starting with the largest site.

¹Chapter 9 describes recent studies suggesting that fusion peptide exposure is affected by inhibition of conformational changes in the head domains.

The target site should also have features conducive to ligand binding. It should be solvent accessible and free of carbohydrate residues. Density for most of the carbohydrate chains is not visible in the electron density maps [1] so the possibility of a site being filled by sugar moieties was estimated by its proximity to glycosylated residues 8A, 22A, 38A, 81A, 165A, 285A, and 154B [1]. Since affinity is determined predominantly by hydrophobic interactions [5], surfaces proposed to interact directly with a ligand should contain contributions from hydrophobic residues.

The six sites in the stem region represented by at least fifteen spheres were evaluated by the above criteria. A ligand bound to cluster 5 may not be able to stabilize the trimer since the site lies only 15 Å from the disordered polypeptide termini at the base of the trimer. Clusters 2 and 3 lie at the interface between monomers and represent parts of interface site 1 discussed below. Cluster 4 lies at the base of a larger site of the hinge region. Since it is distal to the fusion peptide and only 8 Å away from the glycosylated amine of asparagine 38A it was not selected as the highest priority site.

Both clusters 1 and 6 are located in the vicinity of the fusion peptide. Cluster 6 with only 15 spheres represents a site approximately 8 Å deep and 2 Å wide, too small a target for our inhibitor design scheme. In contrast, cluster 1, composed of 153 spheres, encompasses a relatively deep pocket adjoining fusion peptide residues 1B-19B plus several shallower subsites. Large spheres connecting the flat regions to the central site and those spilling out into solvent were eliminated graphically. The resulting

cluster, renamed "tail1A", is shown in Figure 4-1. It contains 54

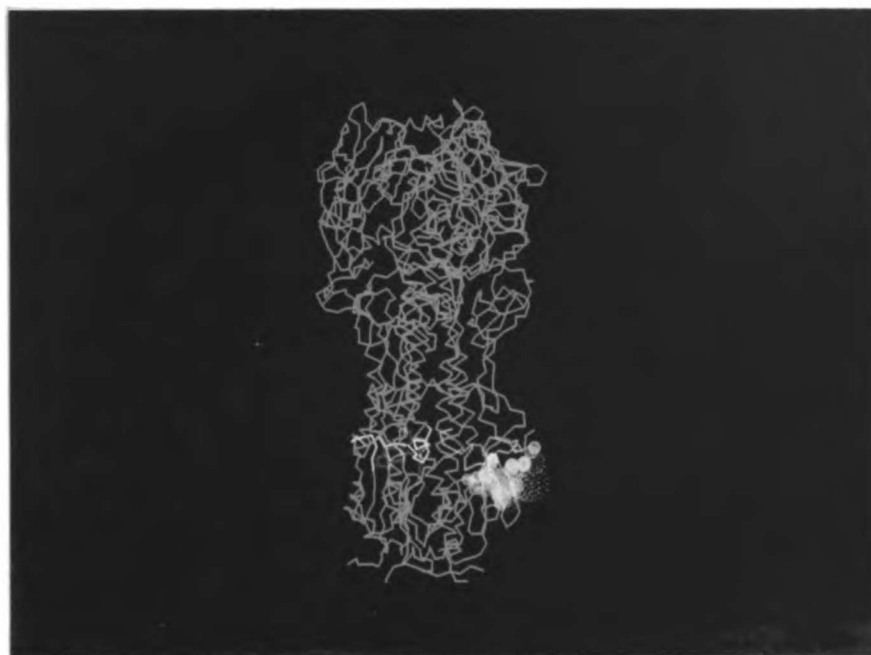


Figure 4-1. Location of the tail1A site on the BHA trimer. Spheres defining the site (yellow) are superimposed on the α -carbon backbone of the trimer (cyan). The three fusion peptides are colored magenta, red, and white.

spheres and lies entirely within the A/B monomer near the interface with the C/D monomer². Residues Lys³ 121D, Arg 124D, Gln 125D lie within 8 Å of the cluster but do not abut directly on the spheres. The site has a bowl-like shape approximately 10 Å wide and 10 Å

²Identical sites exist in each of the monomers.

³Amino acid residues are designated by their three letter abbreviation. Atoms within a particular residue are named following the PDB convention.

deep⁴, and is large enough to contain a small molecule. At the base of the bowl is a dimple-like extension lying over His 17A. The site is solvent accessible, no carbohydrate atoms present in the PDB file lie within 20 Å of the site, and the nearest glycosylated residue, Asn 154D, is 24.5 Å away from the center of the site. Not only does the site lie in the vicinity of the fusion peptide, its core is defined by surface contributed by fusion peptide residues 4B, 7B-19B, and 24B (Figure 4-2).

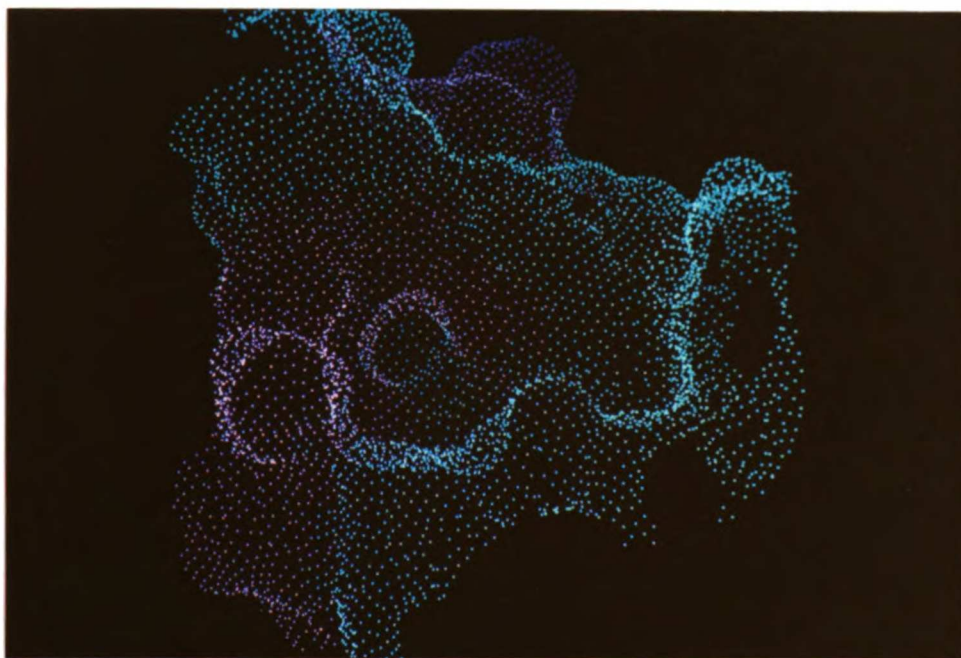


Figure 4-2. Surface of the tail1A site contributed by residues of the fusion peptide. Molecular surface points derived from residues 4B-19B are shown in magenta.

⁴Atom NH1 of Arg 25B is 10.23 Å from atom OE2 of Glu 25B and 11.0 Å from atom ND1 of His 17A.

The tail1A site also has chemical features attractive for inhibitor design (Figure 4-3). Hydrophobic interactions between ligand and receptor determine tight binding while electrostatic interactions contribute to specificity [5]. The site has 111.5 Å² of hydrophobic surface available to interact with potential inhibitors. Small molecules specific for this particular site can be designed with charged groups oriented to complement the unique spacing of charged residues Glu 11B, Glu 325A, and Arg 25B lining the rim of the site and His 17A⁵ at the base of the site.

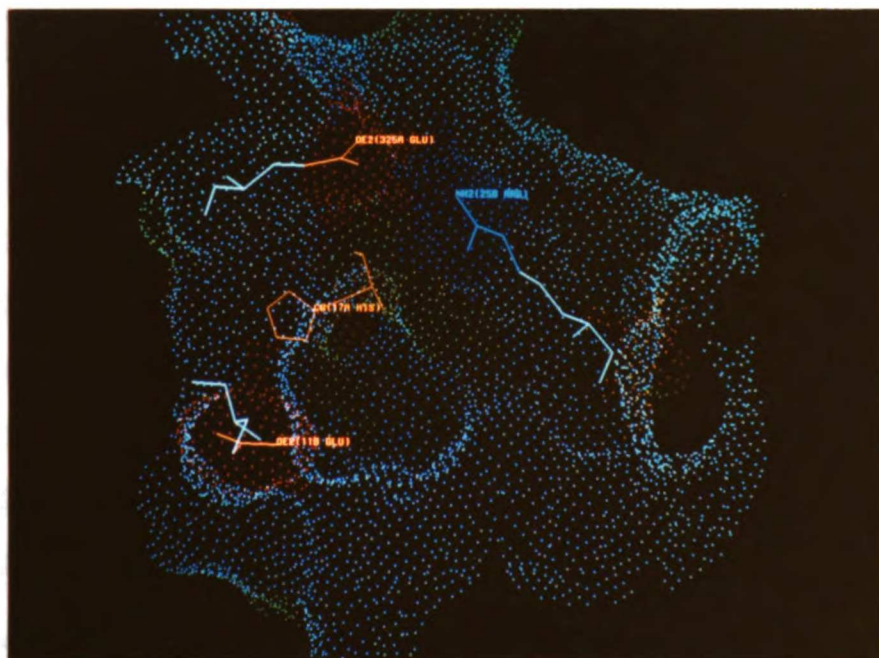


Figure 4-3. The tail1A site colored by chemical functionality. Red represents surface contributed by carboxylic acid moieties of glutamate and aspartate, blue represents the guanidinium of arginine and the amine of lysine. Histidine side chains are colored yellow. Residues Glu 11B, Glu 325A, Arg 25B, and His 17A are labeled.

⁵The charged states of His 17A and His 18A at neutral and endosomal pHs are unknown.

Hydrophobic surface contributed by the side chains of tryptophan, phenylalanine, methionine, leucine, isoleucine, valine, proline, and alanine is shown in green.

ALA 11A	LYS 326A	GLY 16B	GLN 34B
THR 12A	GLN 327A	MET 17B	ALA 35B
LEU 13A	THR 328A	ILE 18B	MET 115B
CYS 14A	GLY 4B	ASP 19B	PHE 119B
LEU 15A	ALA 5B	TRP 21B	GLU 132B
GLY 16A	ILE 6B	TYR 22B	MET 133B
HIS 17A	ALA 7B	GLY 23B	GLY 134B
HIS 18A	GLY 8B	PHE 24B	ASN 135B
ALA 19A	PHE 9B	ARG 25B	GLY 136B
VAL 20A	ILE 10B	HIS 26B	CYS 137B
PRO 21A	GLU 11B	GLN 27B	PHE 138B
MET 320A	ASN 12B	ASN 28B	ILE 149B
VAL 323A	GLY 13B	GLY 31B	LYS 121D
PRO 324A	TRP 14B	THR 32B	ARG 124D
GLU 325A	GLU 15B	GLY 33B	GLN 125D

Table 4-4. Residues of hemagglutinin containing at least one atom within 8 Å of the spheres comprising the tail1A cluster. Distances are computed from the crystallographic coordinates in PDB entry 1hmg.

Finally, the conservation of the site among various strains of influenza was examined (Tables 4-4 and 4-5 and Appendix B). The residues comprising the site are approximately 97% identical among all sequenced H3 hemagglutinins, and 82% identical among all known influenza A hemagglutinins. The high rate of conservation is appealing for two reasons. First, it suggests that this previously unrecognized site may have a role critical for replication of the virus. If so, a compound bound to the site could interfere with this unknown function. Secondly, the high conservation suggests that a compound bound to the site that inhibits X31 influenza will be active against a wide range of strains. However, the conservation among known strains does not guarantee that mutagenic escape would not occur.

sequences	number of sequences HA1/HA2	% identity	
		tail1A site	complete HA
human H3	19/11	99.5	96.3
all H3	40/32	96.7	94.4
all	85/70	81.9	73.6

Table 4-5. Conservation of residues defining the tail1A site among hemagglutinins from all sequenced strains of influenza A. The residues compared are listed in Table 4-4 and Appendix B. Sequence data for several strains is only available for the less highly conserved HA1 polypeptide.

DOCKing the Cambridge Crystallographic Database

Because of its attractive biological and chemical features, the *tail1A* cluster was selected as the target site for the first round of inhibitor design. DOCK1.1 [6] was used in a three step procedure to identify small molecules that could potentially bind to the site. First, databases containing a representative sample of the approximately 80,000 crystallographically determined molecular structures contained in the Cambridge Crystallographic Database (CCD) [7] were docked into the site. High scoring compounds were screened graphically to identify classes of chemical structures with shapes complementary to that of the site, or which positioned substituents in the proximity of one or more of the charged groups surrounding the site. Next, the structural features identified were used to select compounds directly from the unabridged CCD. Compounds in these new databases, treated as molecular shapes only, were then docked into the site. A relatively small number of skeletons which maximized the desired interactions on synthetically accessible frameworks were selected for further consideration. In the third stage, actual chemical functionality was incorporated into the molecular scaffolds in order to satisfy the electrostatic, hydrophobic, and hydrogen bonding requirements of the receptor site. The best designed ligands are either commercially available or readily synthesized, structurally rigid, chemically stable, and easily desolvated. Requirements of the screening assay, such as solubility in aqueous media, were also considered.

Step 1. DOCKing condensed CCD databases

The databases *new.db*, containing 9774 compounds representative of the variety of molecular shapes available in CCD [8], and *dfdat.db*, containing 2745 additional structures, were docked against the site using the parameters listed in Table 4-6. Table 4-7 summarizes the results. In version 1.1 of DOCK, the skeletons are ranked by how well the shape of the compound complements the shape of the site. The exponential scoring function based on the distances between the atoms of the docked ligands and the atoms of the receptor approximates a van der Waals interaction energy.

variable	variable name	value
tolerance for internal distance comparison	<i>dislim</i>	2.0
minimum number of pairs per match	<i>nodlim</i>	8
maximum ratio of ligand and receptor internal distances	<i>ratiom</i>	0.0
close contact limit for scoring	<i>concut</i>	2.3
exponential scoring cutoff	<i>dmin</i>	3.5
maximum distance contributing to score	<i>discut</i>	5.0
maximum number of atoms allowed	<i>natmin</i>	15
minimum number of atoms allowed	<i>natmax</i>	35
number of ligands to save	<i>nsav</i>	500

Table 4-6. Parameter values used in the DOCK1.1 runs. A detailed description of the variables has been published elsewhere.

The 300 ligands with the best shape score from each database were examined graphically to assess the types of skeletons appropriate for the site. Polycyclic ring systems fill the wide bowl of the site well. Five-membered rings lie nicely across the midsection of the site, although their relative flatness prevents them from completely filling the site. Derivatives of both these frameworks have substituents that could interact with the charged residues surrounding the site or with the histidine at the base of the site.

Step 2: DOCKing Specialized Databases

All representatives of the classes of compounds found to be particularly interesting based on the results of the initial runs plus related structures were retrieved from the CCD and docked to the site using the parameters in Table 4-6, unless otherwise noted. The databases included adamantyl compounds (adam), antibiotics (antibiotic), glycosides (glyc), *meta*-benzoic acids (m-benz), purines and pyrimidines (purine, pyr.db, pyr2.db, pyr3.db), octaphanes (octa), bicyclo[2.2.2] compounds (222), and fused 5-membered ring systems (like7). Databases of di- and tripeptides derived from high resolution structures in the PDB [9] were also docked. Although peptides are easily acquired, they were expected to fit relatively poorly due to their linearity. Profiles of the DOCK runs are given in Table 4-7.

database	number of ligands		number of atoms		score ⁶			time (min)
	total	positive	min	max	min	mean	max	
dfdat	2745	436	15	35	53.0	128.0	202.0	393
new.db	9774	4290	15	35	8.3	137.7	123.3	3000
adam	104	104	4	35	47.5	113.3	170.9	34
antibiotic	386	284	10	60	44.3	126.5	217.6	169
glyc	106	84	10	36	79.8	117.6	171.7	28
m-benz	26	26	10	36	64.0	106.2	189.3	5
purine	285	248	10	40	66.6	117.3	200.2	78
pyr.db	456	366	10	40	63.4	113.5	188.9	112
pyr2.db	173	135	10	40	63.5	115.3	189.8	41
pyr3.db	689	579	10	40	29.0	113.2	194.3	147
octa	321	265	10	36	54.5	120.5	189.1	NA
like7	18	15	10	36	43.2	113.6	154.6	4
222	97	60	15	35	90.4	135.0	193.7	48
dipep	1937	1919	10	35	76.9	124.5	208.4	787
tripep	2682	2677	10	35	88.7	136.5	233.2	2035

Table 4-7. Summary of DOCK1.1 runs on excerpted CCD databases. All runs were done on an IRIS 4D/25GT. The classes of structures contained in each database are listed in the text. The total number of ligands in each database is compared to the number of receiving a positive score. Compounds with fewer than the minimum number of atoms or more than the maximum number of atoms were skipped. The minimum, maximum, and mean scores for each database as well as the length of the run in cpu minutes are reported.

The top scoring ligands from all runs were examined graphically for their potential ability to interact with the medial hydrophobic region of the site, with the charged residues, and with His 17A at the base of the site. Compounds that placed groups within hydrogen bonding distance of hydrogen bond donors or acceptors in

⁶The minimum, mean, and maximum scores for the dfdat and adam databases are not indicative of the entire run but only a portion of the database due to a bug in the DOCK1.1 code which has since been remedied.

the receptor were also considered. The skeletons selected for further design are shown in Table 4-8.

Table 4-8. Molecular skeletons from specialized database runs selected for inhibitor design.

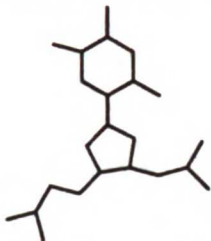
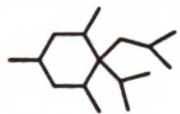
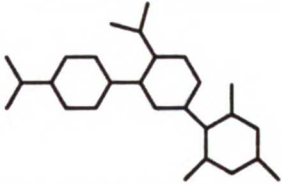
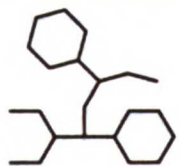
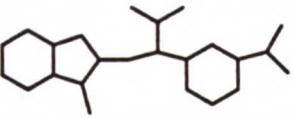
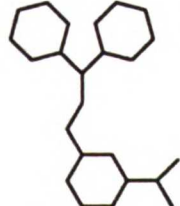
ID	refcode	dock score	structure
1	BULDER	172.5	
2	EADBAC	143.2	
3	MNPHOX	191.8	
4	COFNOA10	191.4	
5	BISYAD	190.1	
6	DOKYEH	170.7	

Table 4-8, continued.

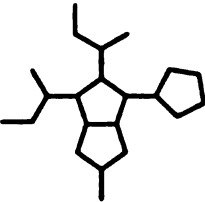
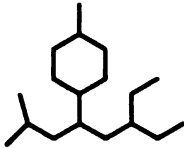
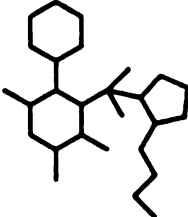
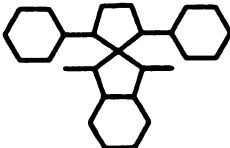
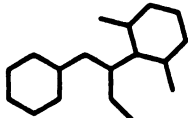
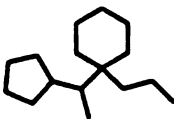
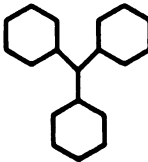
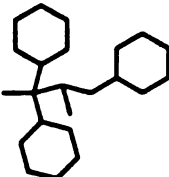
7	BPYRRM	169.6	
8	CYVHZM	158.6	
9	CUVNIQ	157.9	
10	BEJMIM	157.1	
11	PDCBAX	149.9	
12	CUPWOZ	148.7	
13	BIMBUU10	203.4	
14	COLRIE	198.5	

Table 4-8, continued.

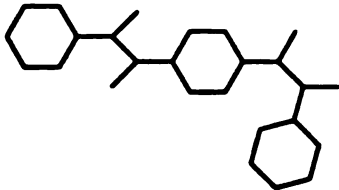
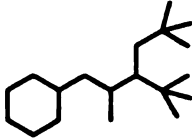
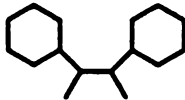
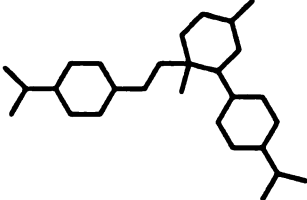
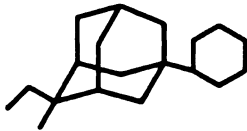

15	CJFEG	180.9	
16	CEYMEY	175.7	
17	CIBHAW	166.8	
18	CIFREO	198.3	
19	ZCEYP10	159.3	
20	CAXNEU	151.2	

Table 4-8 notes. Molecular skeletons from specialized database runs selected for inhibitor design. Refcodes represent the crystallographic entry of CCD from which the frameworks are derived. The DOCK score is proportional to the extent of proposed van der Waals interactions between the site and the ligand in its docked orientation.

Step 3. Chemical Compound Design

With the help of organic chemists, skeletons were converted into actual chemical compounds by incorporation of functional groups in positions maximizing hydrophobic, electrostatic, and hydrogen-bonding interactions with the protein. Emphasis was placed on designing compounds that would be easily synthesized. The two most promising series of leads were the nucleosides (based on structure 1 in Table 4-8) and the adamantyls (based on structures 19 and 20 in Table 4-8). These frameworks both provide opportunities for constructing series of derivatives with a range of functional groups in a variety of different positions for alternate interactions with the site. Commercial derivatives are also available for comparison with the designed structures.

Interestingly, the small molecule with the second highest score in the new.db run was an adamantyl derivative. Amantadine and related compounds are known to have anti-influenza activity at low concentrations, in addition to the lysosomotropic effect of the amine functionality observed at high concentrations [10, 11]. Although the primary target of amantadine at low concentrations is apparently the viral M2 protein [12, 13], there is evidence suggesting that hemagglutinin may also be involved [11, 14, 15].

Regardless of the site of biological action of amantadine, the modeling suggests that the adamantyl group is a good framework for designing compounds to bind to the tail1A site. Dock runs with amantadine revealed a good fit to the wide part of the site (Figure 4-4). Other adamantyl derivatives contained in the CCD were also

docked into the site. The adamantyl groups of these compounds dock primarily to two locations, in the wide bowl of the site and in a groove by His 26B above the wall opposite the fusion peptide (see Figure 4-2). In addition to filling the site well, several of the compounds were positioned with groups reaching down to the deepest part of the pocket, poised for interaction with His 17A.

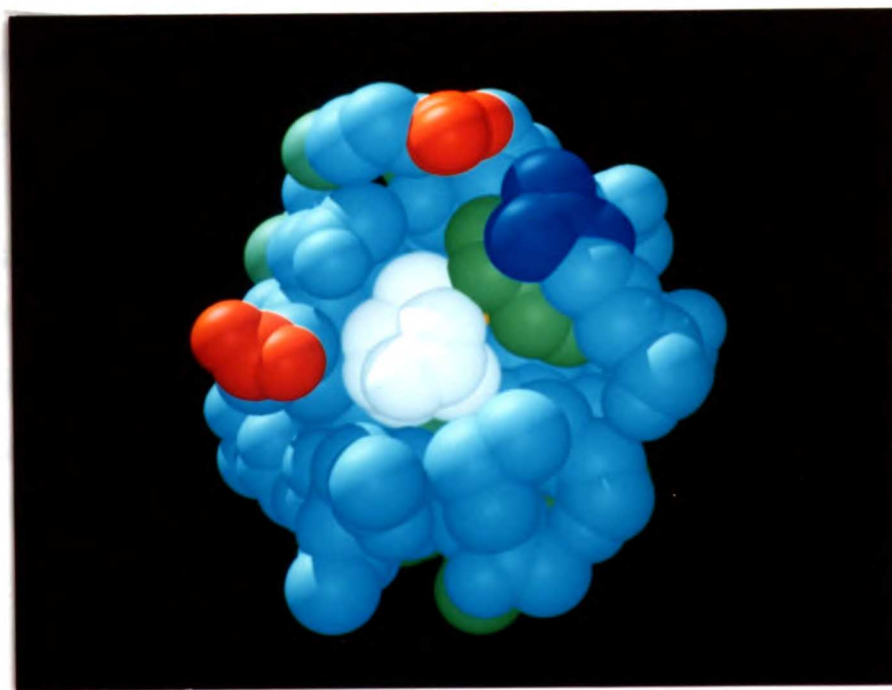


Figure 4-4. Space filling model of amantadine in the tail1A site. The site is colored by chemical functionality as in Figure 4-3. Amantadine is shown in white.

Several derivatives of adamantane were designed to interact with various combinations of the charged groups surrounding the

site (Figure 4-5). If amantadine does bind to the site, it is possible that the additional substituents will increase both the affinity and the specificity of the drug. Even if the tail1A site is not the site of activity of amantadine, the additional substituents may direct it there. Other adamantyl derivatives are commercially available and were used as controls in the screening assays.

The amino group at position 1 of 1-amino-3-(3-aminoadamantyl)-butanoic acid (derivative (a) in Figure 4-5) was predicted to interact with the carboxylate of Glu 11B. The α -amino butyric acid substituent, which is too large to reach into the histidine pocket, was proposed to bind Arg 25B and Glu 325A. Adamantyl derivatives 3-(3-aminoadamantyl)propanoic acid and 3-(3-aminoadamantyl)propanol (derivatives (b) and (c) in Figure 4-5) were designed so that the side chain could reach the histidine pocket. Both the acid and the alcohol forms were proposed since the charge state of His 17A is unknown.

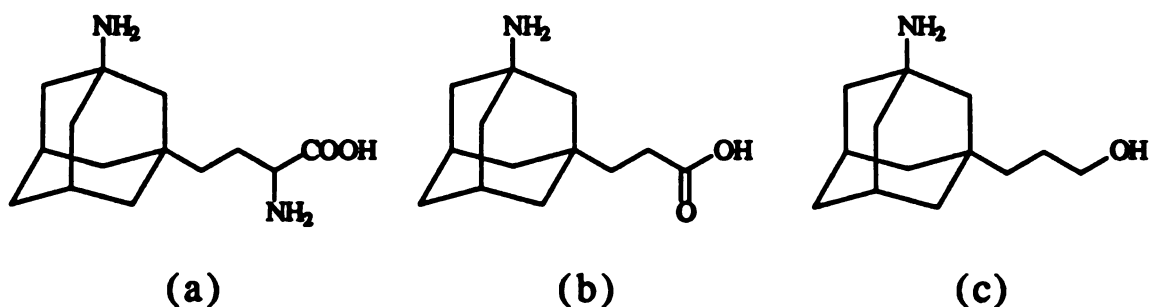


Figure 4-5. Designed adamantyl derivatives. a) 1-amino-3-(3-aminoadamantyl)-butanoic acid. b) 3-(3-aminoadamantyl)propanoic acid. c) 3-(3-aminoadamantyl)propanol.

Nucleosides were the second class of compounds selected as structural frameworks for inhibitor design. Unnatural analogs are frequently used as antiviral agents to inhibit processes dependent on naturally-occurring nucleotides [16]. However, here nucleosides are proposed strictly as frameworks for orienting specific functionalities in a way complementing the requirements of the binding site. Nucleoside derivatives make good stereosynthetic platforms since they may be derivatized in a wide variety of positions. Not only is there an extensive literature on derivatization procedures, but a large number of analogs are commercially available.

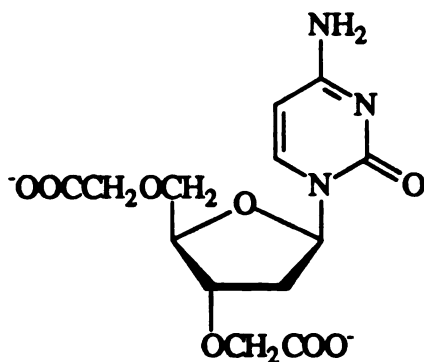


Figure 4-6. 3',5'-O,O-bis-(carboxymethyl)-2'-deoxycytidine, the designed nucleoside with the maximum number of predicted interactions with the site.

In the orientation in which DOCK places the nucleosides, groups of the nitrogenous base are poised for interaction with the charged residues lining the rim of the site while positions on the sugar ring point toward the pocket above His 17A. The compound shown in Figure 4-6 was designed to maximize hydrogen-bonding and charge interactions with the site. 3',5'-O,O-bis-(carboxymethyl)-2'-

deoxycytidine is proposed to interact with all of the acidic and basic groups lining the rim of the protein site except for Glu 325A and to interact with the histidines at the base of the site.

A number of analogs of 3',5'-O,O-bis-(carboxymethyl)-2'-deoxycytidine have been proposed for synthesis or purchase to investigate the roles of the various features. Table 4-9 shows those structures that were actually tested experimentally. These include compounds with variable length chains at the 3' and 5' positions, compounds with the carboxyl groups at the ends of these chains replaced by hydroxyls, and compounds in which the amine functionality of the base has been protected. Cyclic nucleosides have also been investigated since these nucleosides are more rigid than their uncyclized partners. In addition, by rotating the base by 180°, they present a different series of hydrogen bond donors and acceptors to the binding site and provide a wider range of binding determinants in this region of the molecule. AMBER energy minimization was run on some of these derivatives in the context of the binding site with the goal of refining the list of synthetic targets. However, no significant discrimination between the various analogs was observed.

The skeletons in Table 4-8 were also used as the basis of structural searches of chemicals available from Aldrich, Inc. A variety of the compounds resulting from these searches were selected for testing, with particular attention paid to the correct electrostatic properties in order to complement the emphasis DOCK places on spatial complementarity. Compounds which were purchased and experimentally screened are listed in Table 4-9.

Table 4-9. Round I compounds.

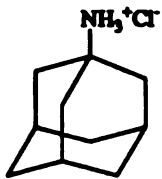
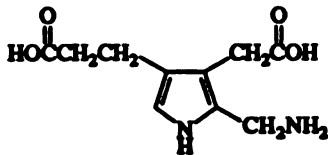
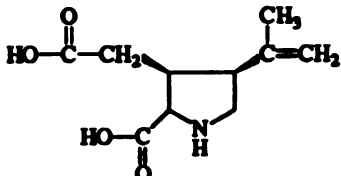
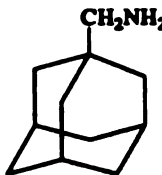
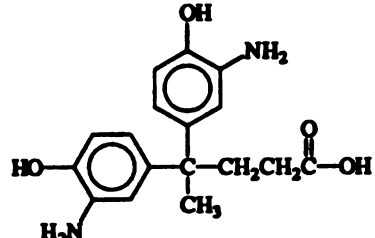
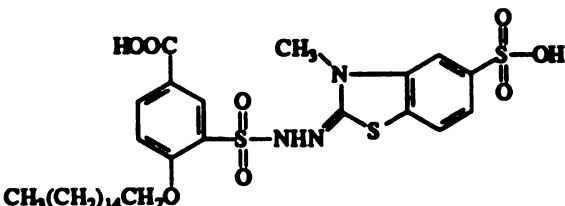
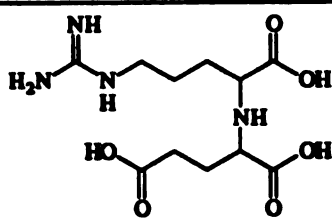
ID	name	source	FW	structure
1	1-adamantanamine hydrochloride	Aldrich	188	
2	porphobilinogen hydrate	Aldrich	226	
3	kainic acid monohydrate	Aldrich	231	
4	1-adamantane methylamine	Aldrich	165	
5	4,4-bis-(3-amino-4-hydroxyphenyl)-valeric acid	Aldrich	316	
6	carboxy-hexadecyloxybenzene sulfonic methyl-sulfo-benzothiazolinyldene hydrazide	Aldrich	684	
7	nopaline monohydrate	Aldrich	322	

Table 4-9, continued.

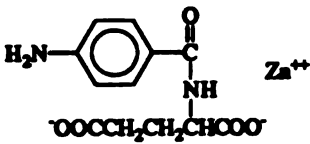
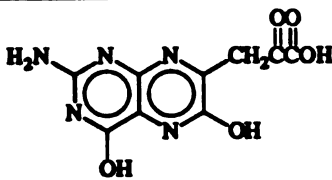
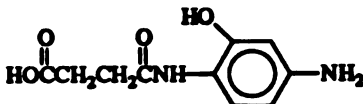
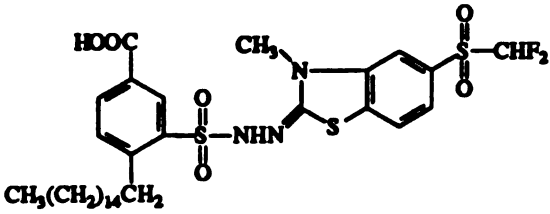

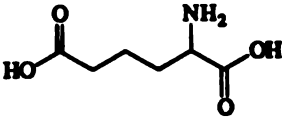
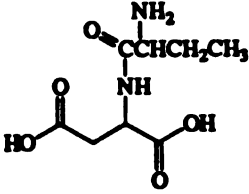
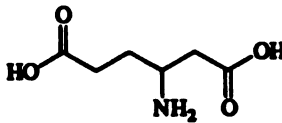
8	N-(4-aminobenzoyl)-glutamic acid, Zn salt	Aldrich	331	 <chem>Nc1ccc(cc1)C(=O)NCC(=O)[O-]</chem> Zn^{++}
9	erythropterin	Aldrich	265	 <chem>Nc1nc2c(ncn2C(O)CC(=O)O)c(O)n1</chem>
10	4'-amino-2'-hydroxysuccinamic acid	Aldrich	224	 <chem>Nc1ccc(cc1NC(=O)CC(=O)O)O</chem>
11	5-carboxy-2-(hexadecyloxy) benzene sulfonic acid-(5-difluoromethyl sulfonyl-3-methyl-2-bromothiazolinyldene) hydrazide	Aldrich	718	 <chem>CCCCCCCCCCCCCCCCOS(=O)(=O)c1ccc(cc1N2C(=S)C(C)C(S(=O)(=O)C(F)F)N2)C(=O)O</chem>
12	1-adamantane carboxylic acid	Aldrich	180	 <chem>OC(=O)C12CC3CC4CC5C(C1)CC2(C3)CC4</chem>
13	DL-2-aminoadipic acid hydrate	Aldrich	161	 <chem>N[C@@H](CCC(=O)O)C(=O)O</chem>
14	L-2-aminobutyryl-L-aspartic acid hydrate	Aldrich	218	 <chem>N[C@@H](CCC(=O)N[C@@H](C(=O)O)C(=O)O)C(=O)O</chem>
15	(+/-)-3-aminoadipic acid	Aldrich	161	 <chem>N[C@@H](CCC(=O)O)C(=O)O</chem>

Table 4-9, continued.

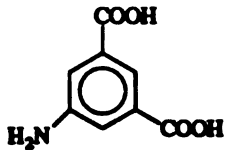
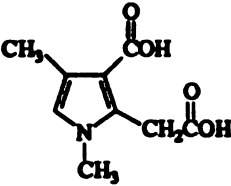
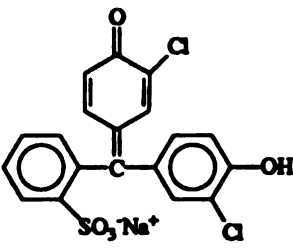
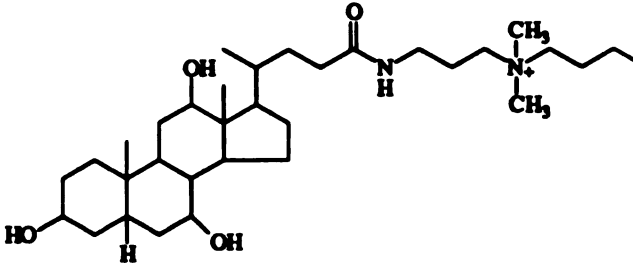
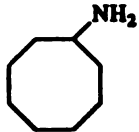
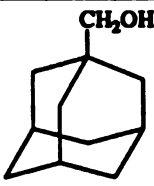
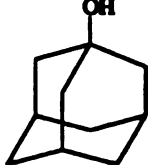
16	5-aminoisophthalic acid	Aldrich	181	
17	3-carboxy-1,4-dimethyl-2-pyrrole acetic acid	Aldrich	197	
18	chlorophenol red, water soluble	Aldrich	445	
19	3-[(3-cholamidopropyl) dimethylammonio]-1-propane-sulfonate dihydrate	Aldrich	651	
20	cyclooctylamine	Aldrich	127	
21	1-adamantane methanol	Aldrich	166	
22	1-adamantanol	Aldrich	152	

Table 4-9, continued.

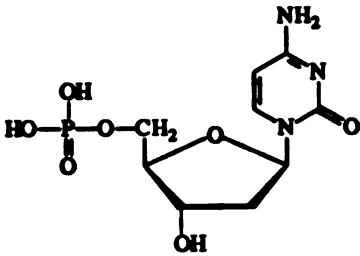
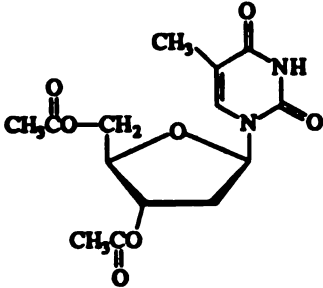
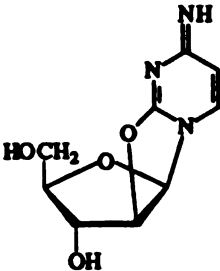
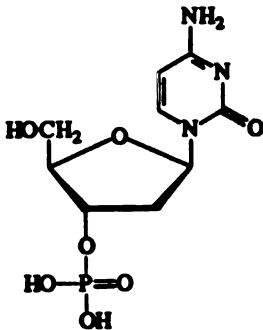
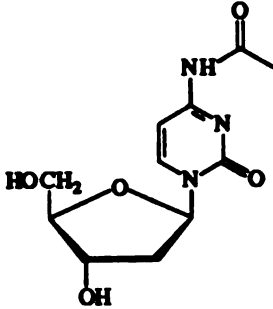
23	2'-deoxycytidine-5-monophosphoric acid monohydrate	Aldrich	325	
24	3',5'-diacetylthymidine	Aldrich	326	
25	(-) cyclocytidine hydrochloride	Aldrich	262	
26	2'-deoxycytidine-3'-monophosphate	Aldrich	307	
27	N ⁴ -acetyl-2'-deoxycytidine	Aldrich	269.	

Table 4-9, continued.

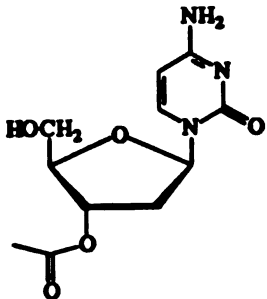
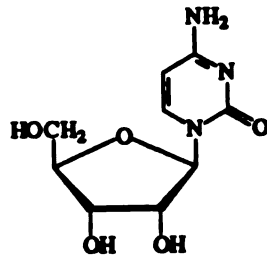
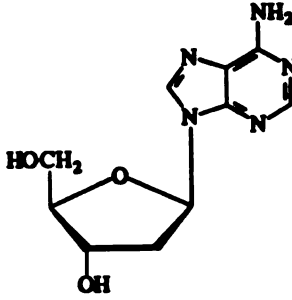
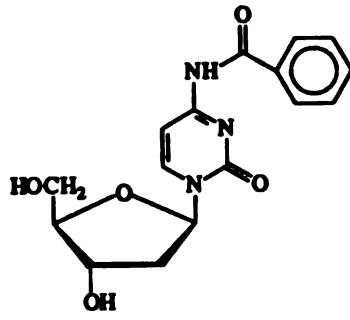
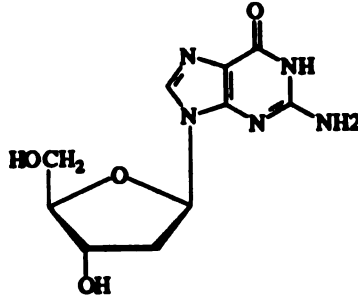
28	3'-O-acetyl-2'-deoxycytidine	Aldrich	269	
29	cytidine	Aldrich	243	
30	2'-deoxyadenosine monohydrate	Aldrich	269	
31	N ⁴ -benzoyl-2'-deoxycytidine	Aldrich	331	
32	2'-deoxyguanosine monohydrate	Aldrich	285	

Table 4-9, continued.

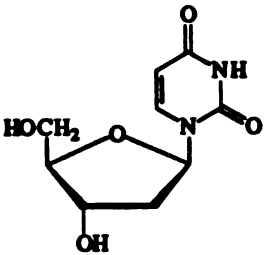
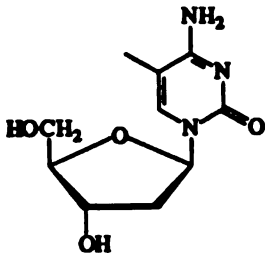
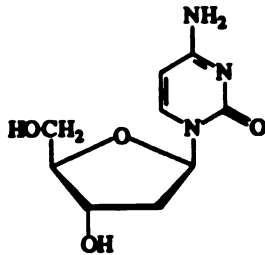
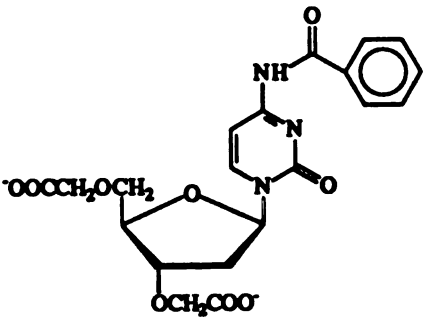
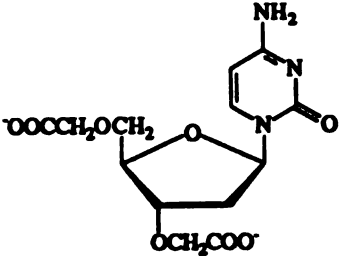
33	(+)-2'-deoxyuridine	Aldrich	227	
34	5-methyl-2'-deoxycytidine	Aldrich	241	
2'αC	2'-deoxycytidine	Sigma	226	
njo6	3',5'-O,O-N ⁴ -benzoyl-2'-deoxycytidine	SERES	445	
njo7	3',5'-O,O-bis-carboxymethyl-2'-deoxycytidine	SERES	340	

Table 4-9, continued.

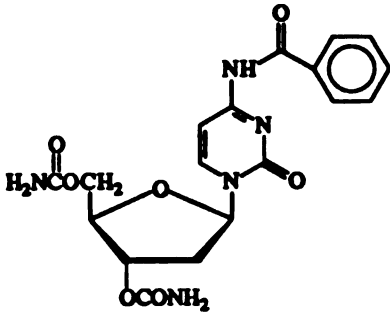
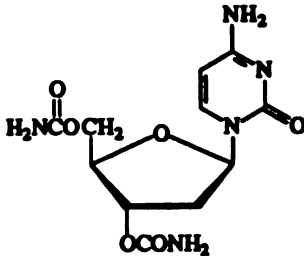
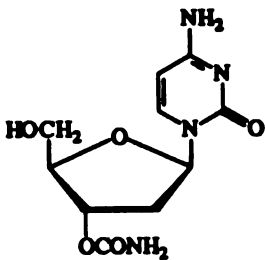
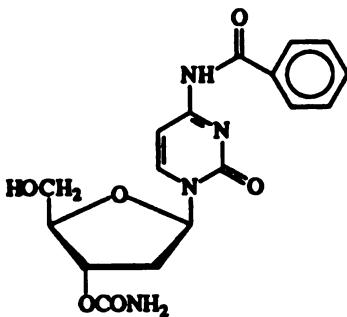
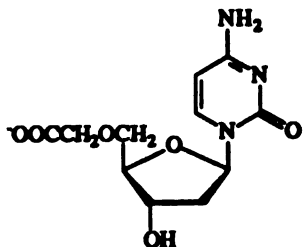
njo8	N ⁴ -benzoyl-3',5'-O,O-biscarbamyl-2'-deoxycytidine	SERES	415	
njo9	2'-deoxycytidine-3',5'-dicarbamate	SERES	310	
njo12	2'-deoxycytidine-3'-carbamate	SERES	267	
njo14	N ⁴ -benzoyl-2'-deoxycytidine-3'-carbamate	SERES	372	
njo26	5'-O-carboxymethyl-2'-deoxycytidine	SERES	282	

Table 4-9, continued.

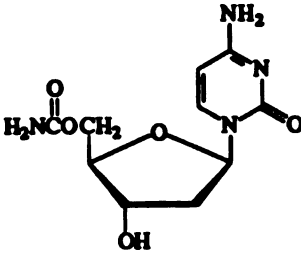
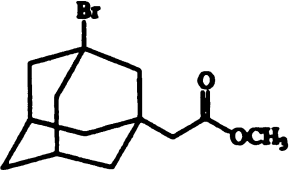
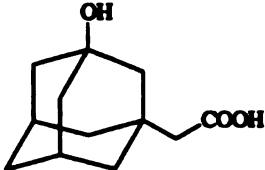
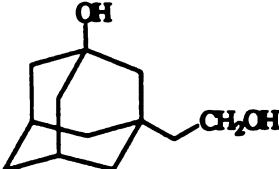
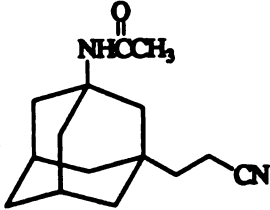
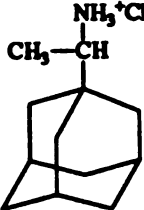
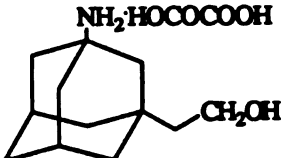
njo27	2'-deoxycytidine-5'-carbamate	SERES	267	
ncco4	methyl-3-bromo-1-adamantane acetate	Aldrich	287	
ncco5	3-carboxymethyl-1-amantanol	SERES	208	
ncco6	3-(2-hydroxyethyl)-1-adamantanol	SERES	194	
ncco7	3-(2-cyanoethyl)-N-acetyl amantadine	SERES	244	
	rimantadine	Hoffman LaRoche	216	
adm1	3-amino-1-adamantaneethanol, oxallic salt	Patricia	285	

Table 4-9, continued.

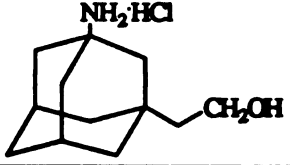
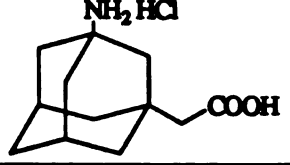
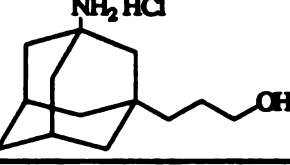
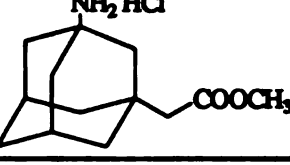
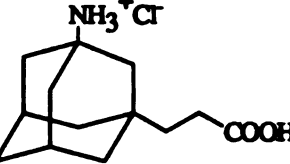
adm2	3-amino-1-adamantaneethanol hydrochloride	Patricia	231	
adm3	3-amino-1-adamantaneacetic acid	Patricia	245	
adm6	3-amino-1-adamantane-(3-propanol)	Patricia	245	
adm7	methyl-3-amino-1-adamantaneacetate	Patricia	259	
adm8	3-amino-1-adamantane propionic acid	Patricia, SERES	260	

Table 4-9 notes. Compounds are identified by name and serial ID number. Formula weights and chemical structures are given for each. Companies where commercial compounds were purchased are listed. Syntheses were done by Patricia Caldera and by Seres, Inc.

Interface sites

The tail1A site was found by analyzing the molecular surface of a single monomer. However, potential binding sites may exist at the interface between two monomers. DMS and SPHGEN were used to find such sites. However, even just the lower third or stem region (see Table 4-1) of the trimer is too large to be used with these programs. Instead, the molecular surface of the regions of the A/B monomer within 20 Å of the C/D monomer and the regions of the C/D monomer within 20 Å of the A/B monomer were selected. Six clusters with at least 15 spheres were identified in this surface. Site 1 lies near residue 109B. It is a large site with three branches, and is lined by residues of the fusion peptide. Site 2 lies by 117D. Part of this site also adjoins on the fusion peptide. Site 3 lies by 153D near the edge of the interface and may be too small for inhibitor design. Site 4 is near 318C at the top of the stem region and may also be too small. Sites 5 and 6 are near residues 171B and 164D, respectively, at the base of the trimer.

Using the criteria for site selection described above, the most promising candidate is site 1. The central SPHGEN site calculated with *radmax* = 5.0 and *radmin* = 1.3 contains 80 spheres. DOCKing new.db to this site using DOCK1.1 with the parameters listed in Table 4-10 revealed several molecular skeletons with good steric fits to the site. The summary of the DOCK run is listed in Table 4-11 and the structures of the selected compounds are shown in Table 4-12. Experimental testing of analogs of these compounds is pending.

parameter	value
<i>dislim</i>	1.5
<i>nodlim</i>	4
<i>ratiom</i>	0
<i>concut</i>	2.3
<i>dmin</i>	3.5
<i>discut</i>	5.0
<i>natmin</i>	10
<i>natmax</i>	36

Table 4-10. Parameter values used to dock the new.db database to stem interface site 1 with DOCK1.1.

number of ligands in database	11943
number assigned positive scores	8509
minimum contact score	30.8
mean contact score	148.0
maximum contact score	296.4
run time (hrs:mins)	209:41

Table 4-11. Output profile from docking the new.db database to stem interface site 1 with DOCK1.1 on an IRIS 4D/25GT.

Table 4-12.

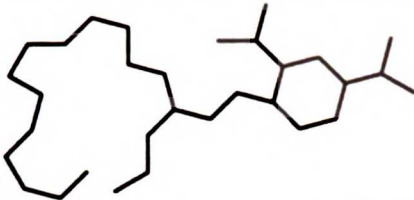
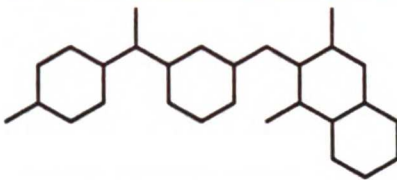
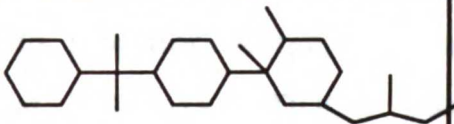


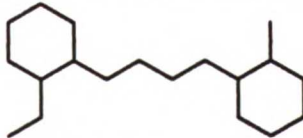
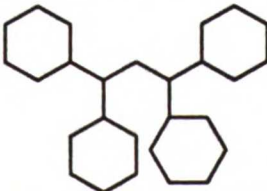
DOCK rank	contact score	refcode	structure
1	296.4	BEMWEV01	
4	267.6	BERGUA	
10	244.0	BOFSUK	
17	239.9	CEBHIA	
28	232.8	BIXLAV	
89	218.7	BMBZDS	
97	217.7	BISRUO	

Table 4-12. Selected ligands from new.db docked to tail interface site 1. Refcodes are the CCD codes. The contact score and rank of each compound are reported.

Summary

Analysis of the crystal structure of hemagglutinin revealed a site in the stem region of the trimer likely to regulate fusion peptide exposure. The site is partially composed of residues of the fusion peptide, has chemical properties attractive to inhibitor design, and is highly conserved among different strains of influenza. DOCK was used to search a database for molecular skeletons with shape complementarity to the site. Adamantyl derivatives and nucleosides were selected as synthetically accessible frameworks, and several derivatives were designed.

A second site adjoining the fusion peptide was found at the interface between monomers. Molecular skeletons with shapes complementary to this site were also identified with DOCK.

References

1. Wilson, I. A., Skehel, J. J. and Wiley, D. C., "Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution." *Nature* 289 366-373 (1981).

2. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. and Weng, J., In F. H. Allen, G. Bergerhoff and R. Seivers (Eds.), Crystallographic Databases - Information Content. Software Systems. Scientific Applications. (Data Commission of the International Union of Crystallography, Cambridge, 1987) pp. 107-132.
3. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., et al., "The Protein Data Bank: A computer-based archival file for macromolecular structures." *J Mol Biol* 112 535-542 (1977).
4. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. and Ferrin, T. E., "A geometric approach to macromolecule-ligand interactions." *J Mol Biol* 161 269-288 (1982).
5. Fersht, A. R., "Basis of biological specificity." *TIBS* 9 145-147 (1984).
6. DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R., "Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure." *J Med Chem* 31 722-729 (1988).
7. Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., et al., "The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information." *Acta cryst B* 35 2331-2339 (1979).
8. Seibel, G. L., personal communication.

9. Meng, E. C., Bodian, D. L. and Kuntz, I. D., personal communication.
10. Hay, A. J. and Zambon, M. C., "Multiple actions of amantadine against influenza." In Y. Becker (Eds.), Antiviral Drugs and Interferon. (Martinus Nijhoff, Boston, 1984) pp. 301-315.
11. Hay, A. J., Zambon, M. C., Wolstenholme, A. J., Skehel, J. J. and Smith, M. H., "Molecular basis of resistance of influenza A viruses to amantadine." *J Antimicrob Chemotherapy* 18 Suppl B 19-29 (1986).
12. Belshe, R. B. and Hay, A. J., "Drug resistance and mechanisms of action on influenza A viruses." *J Respir Disease* 10 S52-S61 (1989).
13. Sugrue, R. J., Bahadur, G., Zambon, M. C., Hall-Smith, M., Douglas, A. R. and Hay, A. J., "Specific structural alteration of the influenza hemagglutinin by amantadine." *EMBO* 9 3469-3476 (1990).
14. Goldman, D. W., Pober, J. S., White, J. and Bayley, H., "Selective labelling of the hydrophobic segments of intrinsic membrane proteins with a lipophilic photogenerated carbene." *Nature* 280 841-843 (1979).
15. Scholtissek, C. and Faulkner, G. P., "Amantadine-resistant and -sensitive influenza A strains and recombinants." *J Gen Virol* 44 807-815 (1979).
16. Hirsch, M. S. and Kaplan, J. C., "Antiviral Agents." In B. N. Fields and D. M. Knipe (Eds.), Virology. (Raven Press, Ltd, New York, 1990) pp. 441-468.

Chapter 5

Experimental Testing of Round I Compounds

The Search for a Positive Control

There are several possible ways to test the Round I compounds for inhibition of the conformational change. A positive control is desirable for selecting the best assay and optimizing the conditions for screening potential anti-fusogenic compounds. The control compound should bind to hemagglutinin noncovalently and inhibit exposure of the fusion peptide without affecting the pH of the solution. A literature search revealed only one possible candidate. Norakin (triperidine hydrochloride, Figure 5-1) was found to inhibit infections by Influenza A/Victoria/3/75 (H3N2) (Victoria) and Influenza A/Fowl Plague Virus/Dobson (Hav1 Neq1) (FPV) with an IC_{50} of approximately 15 μ M [1, 2]. Further studies suggested that norakin may act on hemagglutinin. It inhibits FPV-mediated hemolysis by 60% at 200 μ g/ml (60 μ M) and prevents the pH-driven acquisition of trypsin sensitivity of FPV hemagglutinin at an undisclosed concentration¹ [3]. A resistant strain of A/FPV/Weybridge inherited mutations in the HA gene segment [3, 4]. Since FPV is banned from the United States the ability of Norakin

¹The authors did not show controls demonstrating that the observed inhibition is not due to inhibition of trypsin.

to inhibit the conformational change of hemagglutinin from permitted strains had to be determined. The tests were conducted with Victoria virus, since prior work suggested that this strain may be susceptible to norakin, and with X31 influenza (H3N2), the strain used for the crystallographic studies of hemagglutinin.

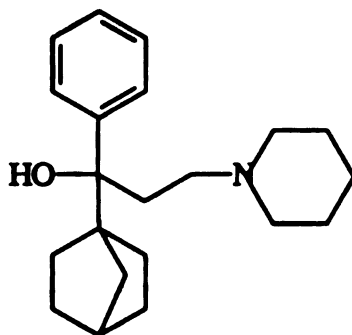


Figure 5-1. Chemical structure of norakin.

Norakin was tested for its ability to inhibit hemolysis by either Victoria or X31 influenza. This assay is described below and the experimental protocol is given in Chapter 3. Significantly increased lysis of the red blood cells (rbc's) was observed in trials with either strain of influenza. Compared to control samples of X31 without drug, norakin at 10^{-4} M caused 15% additional rbc lysis, while 10^{-3} M concentrations doubled the amount of hemoglobin released. The increased lysis was not observed in identical samples maintained at neutral pH, or in low pH-treated control samples containing norakin but not virus. The effect of the drug on hemolysis by Victoria virus is shown in Figure 5-2. As observed with X31, norakin enhanced rbc

lysis in a pH-dependent manner, with complete lysis of rbc's at pH 5.2. The mechanism by which norakin augments rbc lysis is unknown.

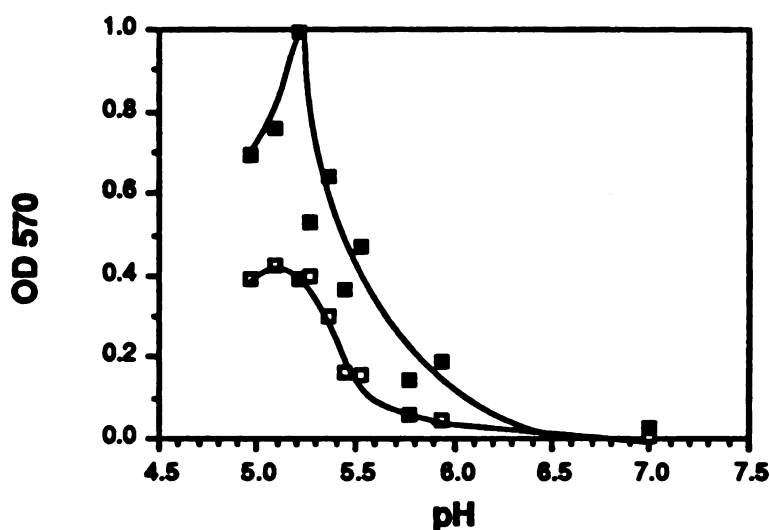


Figure 5-2. Effect of norakin on hemolysis of red blood cells by Victoria virus as a function of pH. Virus, rbc's, and drug were introduced to the appropriate pH for 15 minutes at 37 °C. The concentration of norakin was 0.5 mM (175 µg/ml). Following reneutralization, intact rbc's were pelleted and the OD₅₇₀ of the hemoglobin in the supernatant determined. The experimental procedure is described in detail as hemolysis protocol 1 in Chapter 3. (□) control samples without norakin. (■) samples containing norakin. Standard deviations are approximately 10% of the OD.

Due to the increased lysis, the hemolysis assay could not be used to measure the effect of norakin on the fusion-potentiating changes of hemagglutinin. Instead, the effect of norakin on purified protein was measured by a scintillation proximity assay (SPA). This assay monitors the conformation of ³H-Leu-BHA by testing its ability

to bind conformation-specific antibodies. Chapter 7 describes this assay used in conjunction with an antibody raised against the fusion peptide. Norakin was not able to inhibit precipitation of C22 X31 BHA at mM concentrations. Table 5-1 shows the results of the SPA with Victoria BHA in the presence of 40 μ M norakin, and Figure 5-3 shows the effect at 100 μ M as a function of pH. The results of the SPA under every set of conditions tested showed no significant inhibition of fusion peptide exposure.

pH	norakin	antibody	cpm ppt	% inhibition
7	-	+	51 \pm 10	
7	+	+	62 \pm 13	
7	POST	+	58 \pm 1	
5	-	+	1121 \pm 12	0.0%
5	-	-	64 \pm 11	
5	+	+	1090 \pm 60	2.7%
5	+	-	60 \pm 6	
5	POST	+	1036 \pm 122	7.6%
5	POST	-	61 \pm 8	

Table 5-1. Effect of norakin on the exposure of the fusion peptide of Victoria BHA as measured by SPA. 3 H-Leu-BHA preincubated for 30 minutes in the presence (+) or absence (-) of 50 μ M norakin was introduced to pH 5.0 or maintained at pH 7.0. Following reneutralization, radioactive counts (cpm ppt) were determined by binding to the α -fusion peptide antibody and protein A SPA beads. 50 μ M norakin was added to POST controls between the pH treatment and antibody binding. Inhibition is the % difference in cpm between the norakin-containing samples and the controls without drug.

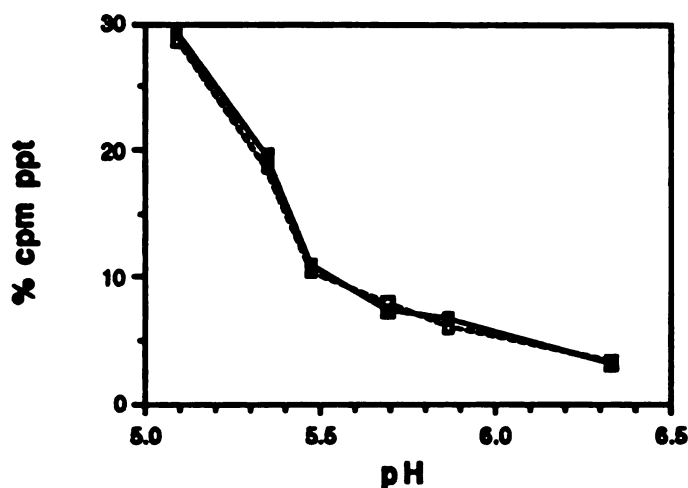


Figure 5-3. The effect of norakin on the conformational change of hemagglutinin as a function of pH. ^3H -Leu-BHA from Victoria virus was incubated for 60 minutes with (■) or without (□) 100 μM drug, then incubated at the appropriate pH for 5 minutes. Radioactive counts were determined by binding to the α -fusion peptide antibody and protein A SPA beads. All precipitations were carried out in duplicate. The percent of total input counts (as determined in scintillation fluid) precipitated is reported.

Testing Round I compounds

Since norakin showed no inhibitory activity with the available strains of influenza, the screening of the Round I compounds had to be done without a positive control. The compounds were tested in a wide variety of assays under a number of different conditions. It was hoped that if any of the compounds did possess inhibitory activity, it would be recognized in at least one of the various experiments. Compounds exhibiting inhibition would then be used for optimization of reaction conditions.

Immunoprecipitation

A polyclonal antibody raised against the fusion peptide is only able to recognize the low pH conformation of hemagglutinin [5]. Compounds which inhibit fusion peptide release will therefore prevent precipitation of hemagglutinin by the α -fusion peptide antibody. Radiolabeled HA or BHA is preincubated with trial ligands, then exposed to low pH. After reneutralization, protein in the low pH conformation is precipitated with the conformation-specific antibodies. Inhibition is detected as a decrease in the number of radioactive counts (cpm) precipitated compared to samples without test ligand. Control samples in which the compound is added after the low pH treatment and reneutralization but prior to antibody binding ("POST") test whether any observed inhibition is simply due to nonspecific interference with the precipitation.

Tables 5-2 and 5-4 list the results of immunoprecipitations with the Round I compounds, and Table 5-3 compares results from some of these compounds to their effects in the SPA. None of the compounds show clear, reproducible inhibition beyond that observed in the POST controls. Compound "norm" did inhibit precipitation by the α -fusion peptide antibody by approximately 20% in some replications of the experiment. Intriguingly, other deoxycytidine derivatives seemed to enhance the conformational changes to a similar extent (Table 5-4). However, the observed activity may simply be due to inaccuracies in the measurements. Determining whether the compounds show consistent behavior in other experiments will distinguish true inhibition from noise.

compound	log conc (M)	n	% inhibition	
			sample	POST
1	-4	14	-3.2 ± 13.3	7.0
	-7	17	1.5 ± 12.2	7.0
	-8	9	4.9 ± 9.3	2.8
2	-4	2	-18.1 ± 8.6	nd
4	-4	6	2.9 ± 5.9	2.3
5	-4	4	57.3 ± 7.2	76.6
6	-4	2	-11.9 ± 7.6	nd
7	-4	2	8.6 ± 10.4	nd
8	-4	6	39.8 ± 9.7	57.4
9	-4	2	-4.7 ± 4.5	nd
10	-4	2	-4.7 ± 12.8	nd
11	-4	2	2.1 ± 2.8	nd
12	-4	2	-7.4 ± 2.8	nd
13	-4	2	-20.9 ± 2.9	nd
14	-4	2	-11.7 ± 10.1	nd
15	-4	4	12.5 ± 5.7	9.9
16	-4	2	-8.4 ± 0.1	nd
17	-4	2	-9.6 ± 1.6	nd
18	-4	6	-2.3 ± 17.1	7.4
20	-4	4	-3.5 ± 4.7	nd
21	-4	4	0.9 ± 1.7	nd
22	-4	2	8.0 ± 6.4	nd
23	-4	2	-3.6 ± 0.1	nd
24	-4	4	2.9 ± 4.9	nd
ncco	-4	6	-5.0 ± 20.6	-2.8
	-7	4	4.2 ± 3.4	-1.0
norm	-4	10	14.4 ± 14.8	6.6

Table 5-2. Effect of Round I compounds on immunoprecipitation of ¹²⁵I-BHA by the α -fusion peptide antibody (see protocol 1 in Chapter 3). Inhibition is the % reduction of cpm precipitated in samples containing test compound compared to control samples without compound. Negative values imply increased cpm. The average and standard deviation of "n" averaged samples are reported. "POST" samples test for nonspecific inhibition of precipitation (see text). "nd", not done.

compound	log conc (M)	% inhibition	
		sample	POST
1	-3	4.3 ± 7.0	nd
2	-3	1.7 ± 0.6	nd
4	-3	22.0 ± 3.5	nd
5	-3	83.5 ± 2.2	nd
6	-3	80.0 ± 0.1	nd
8	-3	80.1 ± 1.2	nd
9	-3	66.6 ± 1.2	50.5
11	-3	19.1 ± 1.3	nd
12	-3	6.8 ± 2.1	nd
13	-3	25.6 ± 4.9	-11.9
18	-3	72.3 ± 4.5	78.8
20	-3	2.6 ± 2.9	nd
22	-3	6.3 ± 2.4	nd
23	-3	16.7 ± 2.8	nd
30	-3	6.5 ± 2.1	nd
31	-3	-6.8 ± 5.9	nd
32	-3	-8.3 ± 4.7	nd
rimantadine	-3	5.2 ± 11.9	nd
adm2	-3	7.5 ± 0.7	nd
adm6	-3	14.4 ± 4.9	nd
adm7	-3	10.4 ± 3.2	nd
adm8	-3	9.1 ± 7.1	nd
ncco5	-3	27.9 ± 1.2	nd
ncco7	-3	0.5 ± 7.4	nd
njo26	-3	10.8 ± 10.0	11.0

Table 5-3. Effect of Round I compounds on ^3H -Leu-BHA in the SPA assay with the α -fusion peptide antibody. Mean values and standard deviations are for duplicate samples except for compounds 18 and njo26 which are averaged over 4 samples. The experimental procedure is described in Chapter 3, except that compounds were preincubated in 10,000 cpm/100 μl BHA in MES Saline. Compounds 5 and 8 were incompletely soluble at mM. Compounds 5, 6, 8, 9, 18 nonspecifically quenched the reactions.

compound	log conc (M)	antibody	% inhibition
2'-deoxycytidine	-4	fp	-46.6 ± 29.0
		C ₂₀	-31.5 ± 26.9
26	-4	fp	-5.6 ± 10.4
		C ₂₀	-9.4 ± 5.6
29	-4	fp	-13.1 ± 3.2
		C ₂₀	-9.0 ± 11.2
33	-4	fp	-8.9 ± 0.9
		C ₂₀	-26.5 ± 0.4

Table 5-4. Effect of selected Round I compounds on immunoprecipitation of ³⁵S-HA. Antibodies against the fusion peptide (fp) and the C-terminus of HA1 (C₂₀) both recognize epitopes in the stem region of hemagglutinin [5]. Experimental details are given in Chapter 3 as immunoprecipitation protocol 1. Inhibition is the % reduction in counts precipitated compared to control samples with test ligand omitted. Mean and standard deviations are reported for duplicate samples, except for 2'-deoxycytidine values which average 4 pairs of duplicates.

Proteinase K Digestion

A second experiment measuring the conformational change of purified hemagglutinin was selected to confirm the results of the antibody binding assays. Low pH hemagglutinin is susceptible to digestion by proteases, whereas the neutral pH form is resistant [6]. The ability of several compounds to inhibit the conversion of hemagglutinin to a proteinase K sensitive form was measured. Purified HA or BHA preincubated with a test compound is exposed to pH 5 then digested with proteinase K. Only undigested protein is

precipitable by trichloroacetic acid [6]. Controls test whether the compounds inhibit the protease rather than the conformational change. The results of this experiment with selected compounds are shown in Table 5-5. None of the compounds showed inhibition at levels above the reproducibility of the experiment.

compound	log conc (M)	% inhibition	number of trials
1	-6	19.1 \pm 6.5	2
	-7	9.8 \pm 12.7	3
	-8	15.8 \pm 11.0	3
	-9	10.1 \pm 11.0	2
	-10	16.1 \pm 19.0	2
25	-4	-18.9	1
26	-4	-12.8 \pm 4.5	2
27	-4	-27.0	1
28	-4	-3.7	1
29	-4	0.8 \pm 9.1	2
33	-4	-4.4 \pm 10.3	2
34	-4	-21.0	1
2'deoxyctidine	-4	-4.7 \pm 14.5	3
njo9	-4	6.6	1
njo12	-4	-8.8	1
njo26	-4	-0.6	1
njo27	-4	6.2	1

Table 5-5. Effect of selected compounds on the proteinase K digestion of radiolabeled HA or BHA averaged over independent experiments. ³⁵S-HA was used for testing all compounds except compound 1 (amantadine) which was tested with ¹²⁵I-BHA. Average reproducibility of duplicate samples in the amantadine trials was 5% of the total number of counts precipitated (range: 1.3% - 8.8%). Reproducibility of the other trials was 3.3% (range: 0.4% - 9.4%). Calculation of % inhibition is described in the methods in Chapter 3.

Aggregation

Another property of the low pH form of hemagglutinin that differentiates it from the neutral pH conformation is its tendency to aggregate in aqueous solution [7]. Aggregation is presumably caused by the increased hydrophobicity of the protein with exposed fusion peptides [7]. Native trimers are separated from aggregated trimers in the low pH conformation by sucrose gradient centrifugation. Sucrose gradients are run under conditions where the 9S trimeric BHA sediments to the center of the gradient and the heavier aggregates sediment to the bottom. Compounds which inhibit the exposure of the fusion peptides should shift BHA-associated radioactivity from the bottom of the gradient to the 9S peak. POST controls in which the compounds are added after the acidification reaction but prior to loading on the gradients test whether the compound is capable of dissociating aggregated trimers.

The feasibility of using this assay for screening compounds was assessed with two ligands (Table 5-6). To determine if the compounds are inhibitory in this assay, the number of counts in the fractions representing the 9S trimer peak are summed and compared to analogous fractions from control samples without compound. However, the trials show that this assay is not reliable enough in a quantitative way to be a useful screen for inhibitors with low activity. The difficulty is twofold: 1) identification of the fractions representing the trimer peak is subjective and 2) the number of cpm in the trimer peak did not always fall to zero in the acid treated control samples.

compound	log conc (M)	% inhibition (trimer peak)
1	- 7	4.7
	- 8	10.7
norm	- 3	0.9
	- 4	1.6 \pm 2.7 n=2
	- 7	21.1 \pm 9.5 n=2

Table 5-6. Aggregation assay results for sample compounds. Total radioactivity in sucrose gradient fractions representing 9S hemagglutinin trimers was determined for reactions with and without test ligands. Calculation of % inhibition is described in the methods.

Of the four assays measuring the conformational change, the SPA gave the most reproducible results both within a single experiment and in independent repetitions. This experiment will be used for testing any additional compounds for inhibition of fusion peptide exposure.

Hemolysis

The ability of the compounds to inhibit hemagglutinin-mediated fusion was tested by hemolysis. This assay measures the extent of fusion between whole virus and intact red blood cells. The amount of fusion is proportional to the amount of hemoglobin

released into the medium [8, 9]. Two sets of controls were done in parallel. As with norakin, the compounds may lyse rbc's whether or not they inhibit the conformational change of hemagglutinin. Alternatively, they may prevent the lysis normally observed when rbc's are introduced to low pH. Control reactions containing the trial ligand but no virus test for these nonspecific effects. Secondly, hemolysis may be inhibited by preventing the virus from binding to its receptors on the rbc's. Hemagglutination assays measure whether any inhibitory activity can be accounted for by this mechanism.

The results of hemolysis for the Round I compounds are shown Table 5-7. None of the compounds gave significant amounts of inhibition. However, amantadine (compound 1) did prevent hemolysis at levels not accounted for by the noise level of the experiment. The exact concentration at which maximum inhibition was observed varied between 0.1 - 10 nM in independent replications, which increased the standard deviations significantly when averaged over multiple experiments. Interestingly, the unusual concentration profile (Figure 5-4) has also been observed in studying the effects of amantadine on the replication of influenza virus [10]. It is unknown whether the observed inhibition of hemolysis is due to inhibition of the conformational change, inhibition of binding (the maximum effect of ~20% inhibition is too small to be measured by hemagglutination), or effects on membrane properties.

Table 5-7. Round I compounds in hemolysis.

compound	log conc (M)	% inhibition	n	lysis
1	- 4	-3.2 ± 4.9	3	
	- 5	3.7 ± 3.3	3	
	- 6	8.9 ± 5.8	5	-
	- 7	11.6 ± 6.4	11	
	- 8	17.7 ± 10.4	12	
	- 9	17.8 ± 10.1	10	-
	-10	13.8 ± 9.2	5	
	-11	7.0 ± 6.2	3	
rimantadine	- 4	12.9 ± 24.7	1	
	- 9	0.0 ± 2.3	1	-
adm1	- 4	-5.3 ± 10.8	5	
	- 5	8.5 ± 12.7	4	
	- 6	6.8 ± 9.0	5	
	- 7	6.9 ± 17.3	4	
	- 8	13.1 ± 18.4	4	
	- 9	8.6 ± 19.1	3	↓
	-10	6.7 ± 1.6	3	↓
	-11	-1.8 ± 3.5	3	
adm2	- 9	-10.0 ± 2.2	1	-
adm3	- 9	-1.4 ± 6.5	1	↓
adm6	- 9	8.3 ± 2.7	1	↓
adm7	- 9	-10.0 ± 5.1	1	-
norm	- 3	18.9 ± 21.6	7	
	- 4	-3.7 ± 7.3	6	
	- 5	-6.4 ± 11.1	6	
	- 6	0.5 ± 12.2	6	
	- 7	-3.9 ± 15.8	6	
	- 8	-5.3 ± 11.4	4	
25	- 4	-2.2 ± 6.0	1	
26	- 4	6.78 ± 12.0	4	-
27	- 4	5.3 ± 5.1	2	-
28	- 4	-6.0 ± 3.1	1	
29	- 4	-0.70 ± 7.8	2	
33	- 4	-1.1 ± 1.6	2	
34	- 4	-7.35 ± 11.4	2	
2'deoxycytidine	- 4	11.4 ± 12.6	8	-

Table 5-7, continued.

njo9	- 4	7.8 ± 14.3	3	-
	- 6	10.5 ± 13.0	2	
njo12	- 4	-7.3 ± 11.8	1	
njo14	- 4	8.0 ± 2.7	1	
njo26	- 4	2.8 ± 2.8	1	
njo27	- 4	1.3 ± 8.0	1	

Table 5-7 notes. All compounds were tested using protocol 1 except rimantadine, adm2, adm3, adm6, adm7, 25, 28, njo12, njo14, njo26, and njo27 which used protocol 2. Compounds 26, 27, 29, 33, 34, and njo9 were tested both ways. % inhibition is the % reduction in OD₅₇₀ compared to control samples without compound. Samples with higher OD than the controls are represented by negative inhibition. The average and standard deviation are reported with the number of independent experiments done (n). Samples within each experiment were carried out in duplicate. Compounds which decreased lysis in the absence of virus are represented by '↓'; '-' means no difference was observed compared to controls.

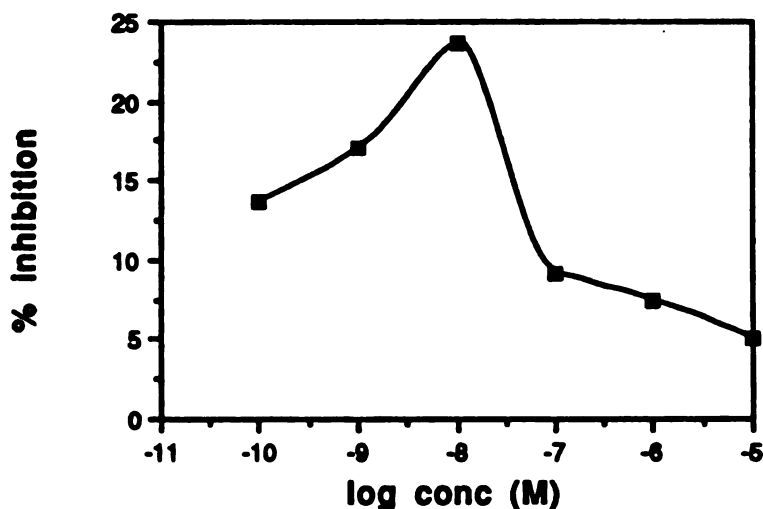


Figure 5-4. Concentration profile of amantadine in a single hemolysis experiment. Standard deviations are 5-10%.

Plaque Assays

Betty Fraser-Smith of Syntex kindly conducted plaque assays on some of the compounds. The results are shown in Table 5-8. None of the nucleoside samples tested had antiviral activity beyond cell toxicity against Influenza A/Japan (H2) or Influenza B/Lee. Of the adamantyl compounds tested, only amantadine and rimantadine showed activity against Influenza A in the μM range. Amantanol (compound 22) did possess some activity as well, with an IC_{50} of 1 mM. Ribavarin was used as a positive control. Syntex samples of the free base and hydrochloride salts of amantadine were included as an indication of the reproducibility of the experiment.

Compound	IC_{50} (μM)		Cell Toxicity (μM)	
	A/Japan	B/Lee	Partial	Complete
norakin	>32	>32	100	320
2	>1000	>1000	--	--
22	1000	>1000	--	--
23	>1000	>1000	--	--
26	>1000	>1000	--	--
2'deoxycytidine	>1000	>1000	--	--
1 (amantadine)	32	>32	100	--
rimantadine	<1	>32	100	1000
adm2	>320	>320	1000	--
adm3	>320	>320	1000	--
adm7	>100	>100	320	--
adm6	>100	>100	320	--
ncco	>1000	>1000	--	--
ribavirin	>100/<320	>32/<100	--	--
amantadine (free base)	100	>320	1000	--
amantadine HCl (Syntex stock)	100	>100	320	

Table 5-8. Results of plaque assays against selected Round I compounds. IC_{50} is the concentration of drug which reduced plaques

by 50%. Cell toxicity is listed as the lowest concentration at which nonviability was observed.

Infectivity

The ability of the Round I compounds to inhibit viral replication was also tested by an ELISA-based infectivity assay. This experiment measures infection by monitoring the appearance of viral hemagglutinin on the surface of the MDCK2 cells. Cells are infected in the presence of compound, then treated with an anti-hemagglutinin monoclonal antibody. A secondary antibody coupled to peroxidase is then bound to the primary antibody. Infection is quantitated as the amount of color produced by a peroxidase-catalyzed reaction. Compounds which prevent influenza infections lead to a decrease in the amount of color produced. Cell viability assays are conducted in parallel to determine whether any apparent antiviral activity is due to cell toxicity.

The Round I compounds were screened by the infectivity and viability assays. Table 5-9 lists the results. Compounds 1, 2, 9, 20, njo9, njo26, njo7, adm2, adm7, adm8 all inhibit infectivity by a greater percentage than viability (where done). This effect may be due to the amine functionality present in each of these structures, since amines are well-known for preventing viral entry by increasing endosomal pH. Interestingly, in some repetitions of the experiment, amantadine (compound 1) was able to inhibit viral replication at 10^{-9} M, in agreement with its effect in hemolysis.

Compounds 11 (at 10^{-4} M), 18, 22, 31, ncco7 all exhibited antiviral activity at high concentrations that is not explicable by cell toxicity, as measured by the MTT cell viability assay.

Table 5-9. Effect of Round I compounds on infectivity.

compound	log conc (M)	n	% inhibition	
			infectivity	viability
1	-3	15	65.4 \pm 30.8	-2.6
	-4	9	2.7 \pm 9.6	-5.4
	-5	9	-3.1 \pm 16.2	nd
	-6	8	-0.3 \pm 5.1	nd
	-7	10	-2.8 \pm 10.6	nd
	-8	15	2.8 \pm 9.4	nd
	-9	15	16.2 \pm 25.5	nd
	-10	15	4.7 \pm 18.0	nd
2	-3	1	45.7	5.0
	-4	1	10.5	1.4
	-5	1	-5.7	nd
	-6	1	-0.1	nd
4	-3	1	108.5	99.5
	-4	1	62.3	52.4
	-5	1	5.6	1.3
	-6	1	-20.7	nd
5	-3	1	102.9	101
	-4	1	-0.4	12.7
	-5	1	4.0	nd
	-6	1	5.1	nd
6	-3	1	89.4	97.1
	-4	1	100.2	97.1
	-5	1	87.5	nd
	-6	1	3.7	nd
7	-3	1	3.6	nd
	-4	1	1.7	nd
	-5	1	4.1	nd
	-6	1	4.4	nd
8	-3	1	107.4	101
	-4	1	13.2	57.7
	-5	1	2.3	nd
	-6	1	4.7	nd
9	-3	1	33.8	7.6
	-4	1	5.9	3.3
	-5	1	5.8	nd
	-6	1	6.0	nd

Table 5-9, continued.

10	-3	1	7.5	nd
	-4	1	0.0	nd
	-5	1	-3.5	nd
	-6	1	-4.3	nd
11	-3	1	108.8	96.6
	-4	1	95.2	35.3
	-5	1	2.7	nd
	-6	1	0.4	nd
12	-3	2	20.5 ± 21.6	35.5
	-4	2	11.6 ± 3.7	7.4
	-5	2	0.7 ± 5.5	nd
	-6	2	2.0 ± 1.6	nd
	-7	1	-0.9	nd
	-8	1	-3.4	nd
	-9	1	2.3	nd
	-10	1	-10.3	nd
	-11	1	5.2	nd
13	-3	1	19.3	20.7
	-4	1	0.7	15.1
	-5	1	3.3	nd
	-6	1	-1.3	nd
14	-3	1	4.2	nd
	-4	1	-1.2	nd
	-5	1	2.5	nd
	-6	1	-5.9	nd
15	-3	1	4.2	19.1
	-4	1	2.1	-2.2
	-5	1	1.2	nd
	-6	1	3.3	nd
16	-3	1	-6.2	nd
	-4	1	1.5	nd
	-5	1	4.7	nd
	-6	1	-0.7	nd
17	-3	2	0.6 ± 5.5	nd
	-4	2	-5.8 ± 4.7	nd
	-5	2	6.3 ± 2.6	nd
	-6	2	-2.8 ± 1.0	nd
18	-3	3	100.2 ± 8.1	101
	-4	3	92.0 ± 10.8	21.8
	-5	3	-4.2 ± 2.4	-7.6
	-6	3	7.6 ± 6.3	nd
	-7	2	0.0 ± 0.0	nd
20	-3	1	82.0	22.7
	-4	1	16.2	10.1
	-5	1	10.9	nd
	-6	1	1.8	nd
21	-3	1	-3.5	4.5
	-4	1	0.0	1.9
	-5	1	2.9	nd
	-6	1	0.4	nd

Table 5-9, continued.

22	-3	1	25.5	8
	-4	1	6.7	1.3
	-5	1	7.8	nd
	-6	1	7.8	nd
	-7	1	-6.3	nd
	-8	1	0.2	nd
	-9	1	0.3	nd
	-10	1	5.2	nd
	-11	1	-7.7	nd
23	-3	1	6.2	nd
	-4	1	0.7	nd
	-5	1	-1.6	nd
	-6	1	3.1	nd
24	-3	1	11.1	-12.5
	-4	1	9.5	-12.8
	-5	1	13.5	-10.7
	-6	1	5.9	nd
25	-3	1	-2.0	nd
	-4	1	-14.0	nd
	-5	1	12.3	nd
	-6	1	12.7	nd
26	-3	1	6.4	nd
	-4	1	-11.3	nd
	-5	1	2.5	nd
	-6	1	5.8	nd
	-7	1	1.4	nd
	-8	1	1.2	nd
	-9	1	0.0	nd
28	-3	1	3.0	nd
	-4	1	6.3	nd
	-5	1	-6.1	nd
	-6	1	-6.7	nd
29	-3	1	5.8	nd
	-4	1	4.3	nd
	-5	1	14.7	nd
	-6	1	14.4	nd
30	-3	1	40.4	88.9
	-4	1	8.9	6.3
	-5	1	0.4	nd
	-6	1	-2.6	nd
31	-3	1	28.8	15.9
	-4	1	11.0	-1.0
	-5	1	9.6	nd
	-6	1	2.5	nd
32	-3	1	29.4	48.4
	-4	1	4.5	4.5
	-5	1	7.0	nd
	-6	1	2.2	nd

Table 5-9, continued.

33	-3	1	12.3	nd
	-4	1	20.3	nd
	-5	1	18.7	nd
	-6	1	24.7	nd
34	-3	1	-17.6	nd
	-4	1	-14.1	nd
	-5	1	-1.1	nd
	-6	1	-14.8	nd
2'deoxy- cytidine	-3	2	-6.3 ± 15.4	nd
	-4	2	9.2 ± 36.9	nd
	-5	2	2.1 ± 34.3	nd
	-6	2	-6.7 ± 8.3	nd
njo7	-3	1	13.6	nd
	-4	1	-2.1	nd
	-5	1	1.1	nd
	-6	1	-3.0	nd
njo9	-3	2	29.7 ± 33.4	nd
	-4	2	8.0 ± 9.4	nd
	-5	2	22.1 ± 26.2	nd
	-6	2	5.0 ± 1.1	nd
njo12	-3	1	-30.4	nd
	-4	1	-5.1	nd
	-5	1	1.9	nd
	-6	1	-4.3	nd
njo14	-4.3	1	-7.4	nd
	-5.3	1	8.3	nd
	-6.3	1	5.6	nd
	-7.3	1	5.1	nd
njo26	-3	2	38.7 ± 16.3	3.7
	-4	2	0.6 ± 37.2	-1.3
	-5	2	-0.6 ± 38.9	3.1
	-6	2	-11.3 ± 27.8	nd
njo27	-3	1	14.9	nd
	-4	1	9.6	nd
	-5	1	11.2	nd
	-6	1	1.8	nd
adm1	-3	3	68.8 ± 42.6	74.5
	-4	3	45.9 ± 28.1	6.7
	-5	3	0.6 ± 4.8	5.0
	-6	3	-0.5 ± 1.8	3.3
	-7	3	1.1 ± 2.2	nd
	-8	3	0.7 ± 6.0	nd
	-9	3	7.4 ± 1.0	nd
	-10	3	-1.7 ± 8.8	nd
	-11	2	0.3 ± 0.4	nd
	-12	2	3.4 ± 6.0	nd

Table 5-9, continued.

adm2	-3	1	67.5	1.4
	-4	1	16.2	-4.4
	-5	1	7.4	nd
	-6	1	18.0	nd
	-7	1	-3.3	nd
	-8	1	15.0	nd
	-9	1	4.4	nd
	-10	1	7.0	nd
	-11	1	9.6	nd
	-12	1	-7.8	nd
adm3	-4	1	2.5	nd
	-5	1	-3.1	nd
	-6	1	7.2	nd
	-7	1	1.5	nd
	-8	1	8.9	nd
	-9	1	-2.6	nd
	-10	1	4.7	nd
adm6	-3	2	59.8 ± 24.7	73.2
	-4	2	-0.2 ± 0.5	8.5
	-5	2	-0.05 ± 1.5	nd
	-6	2	3.5 ± 6.5	nd
	-7	2	-1.1 ± 1.8	nd
	-8	2	-3.8 ± 3.2	nd
	-9	2	1.1 ± 15.8	nd
	-10	2	-3.1 ± 3.7	nd
	-11	1	1.0	nd
adm7	-3	2	57.7 ± 4.0	29.6
	-4	2	3.1 ± 5.9	6.9
	-5	2	0.8 ± 2.8	nd
	-6	2	8.9 ± 13.2	nd
	-7	2	10.8 ± 12.8	nd
	-8	2	4.8 ± 14.3	nd
	-9	2	0.7 ± 10.5	nd
	-10	2	0.9 ± 7.4	nd
	-11	2	-3.0 ± 12.7	nd
adm8	-3	1	25.9	5.5
	-4	1	-4.3	-0.9
	-5	1	8.6	nd
	-6	1	3.1	nd
	-7	1	-3.9	nd
	-8	1	11.0	nd
	-9	1	-4.6	nd
	-10	1	-1.3	nd
	-11	1	3.6	nd
	-12	1	2.0	nd
ncco5	-3	1	10.4	10.3
	-4	1	1.7	4.7
	-5	1	-0.2	nd
	-6	1	-0.7	nd

Table 5-9, continued.

ncco7	-3	1	36.6	5.2
	-4	1	3.8	1.3
	-5	1	-1.5	nd
	-6	1	-1.0	nd

Table 5-9 notes. Effect of round I compounds in the EIA infectivity assay. Each concentration was tested in triplicate in every repetition of the experiment. Inhibition is the % reduction in OD₄₁₀ compared to control values with compound omitted. The average, standard deviation, and number of experiments (n) are reported. Cell viability measured by the MTT assay was determined in parallel. The experimental protocols and calculation of % inhibition for both assays are described in Chapter 3.

Binding Assays

The Round I compounds were designed to bind to the target site. However, the assays described above all measure biological activity. The compounds may bind to the site but not exert any effect on the activity of hemagglutinin. To determine if this is indeed the case, a direct binding assay is required.

Two assays, equilibrium dialysis and gel filtration chromatography, were tested for their feasibility in screening compounds. The major drawback of these assays is that they require the ligands to be radiolabeled to high enough specific activity to detect the possibly weak binding affinity. Only two chemicals structurally related to any Round I compound were attainable in radiolabeled form, ³H-amantadine and 5'-³H-2'-deoxycytidine (2'dC). Even if these particular compounds do not bind to hemagglutinin, this exercise will determine if these binding assays are sufficiently

reliable to pursue acquisition of labeled derivatives of other compounds.

The first assay tested was equilibrium dialysis. In this experiment, a mixture of protein and radiolabeled ligand is incubated in one half of a dialysis chamber, separated by a semipermeable membrane from the other half. The membrane allows the ligand to pass freely between the two chambers, but restricts the protein to one side only. At equilibrium, the concentration of free ligand will be the same on both sides of the membrane. The total amount of radioligand in the well containing the receptor will be increased by an amount related to the binding affinity. If the specific activity of the ligand is known, the difference in the number of cpm on the two sides of the membrane gives the concentration of protein-ligand complexes, which, given the total concentration of protein and ligand, reveals the binding constant.

Binding constant determination requires an accurate knowledge of the concentrations of the ligand and the protein. The concentration of the ligands was determined by the specific activity provided by the manufacturers. The protein concentration was determined in two ways. A Lowry assay revealed the concentration of protein to be $0.813 \mu\text{g}/\mu\text{l}$. Based on the sequence of X31 BHA, the molecular weight of the peptide chain is $1.688 \times 10^5 \text{ g/mol}$ trimer. This gives a concentration of $4.8 \times 10^{-6} \text{ M}$. The concentration was confirmed by taking the UV spectrum of BHA and applying the equation determined by Ruigrok, et al. [11]. By this method, the concentration of BHA trimers was $4.8 \times 10^{-6} \text{ M}$.

For equilibrium dialysis, each cell was loaded with 50 μ l of 5×10^{-7} M 2'dC or 50 μ l of 5×10^{-7} M 2'dC and 10^{-7} M BHA trimers. At equilibrium, the difference in cpm between the halves of eight identical wells was 1047 ± 1543 , or 0.7% of the total input counts. Since no differential was seen under these conditions, the binding constant must be greater than 10^{-4} M. The experiment was repeated with 5×10^{-6} M BHA and 10^{-5} M 2'dC, and again the differential cpm was 0.7% of the total number of counts present. Therefore, equilibrium dialysis measurements imply that 2'dC does not bind to BHA under these conditions.

The second assay attempted was sephadex gel chromatography. Radiolabeled ligands are run through a sepharose column to determine their elution profile. A mixture of the ligand and cold protein is then applied. If binding has occurred the rate of elution will be reduced. 3 H-amantadine was mixed with protein at a final concentration of 8×10^{-6} M ligand and 4.8×10^{-6} M unlabeled BHA. For comparison, 3 H-2'dC was mixed at 5×10^{-7} M with 10^{-7} M BHA. The ligand-protein mixtures are incubated for at least 1 hour at RT, applied to the top of the column, and fractions were collected and counted. No detectable binding was observed with either ligand.

The chromatography assay is not appropriate for systems in which the off-rate is faster than the elution rate. For column lengths minimizing retention times, the concentrations of protein and ligand required to measure low affinity binding produced a smear rather than crisp peaks. Of the two binding assays, equilibrium dialysis was more reliable and highly reproducible. However, due to the difficulty and expense in acquiring radiolabeled ligands and the

limited amount of protein available, this experiment will only be repeated for compounds showing activity in at least one other assay.

Summary

The Round I compounds were tested in assays measuring the conformational change of hemagglutinin, virus-induced hemolysis, and influenza infectivity. Amantadine showed a small level of inhibition in each of these assays, but no reproducible inhibition was detected with any other compound. Of the assays measuring the conformational change, the SPA was the most convenient and reproducible and will be used as the primary assay for screening any additional ligands. The equilibrium dialysis assay measuring binding of radiolabeled ligands to hemagglutinin will be carried out for any compounds obtainable with high enough specific activity.

References

1. Presber, H. W., Schroeder, C., Hegenscheid, B., Heider, H., Reefschlager, J. and Rosenthal, H. A., "Antiviral activity of norakin (triperiden) and related anticholinergic antiparkinsonism drugs." *Acta virol* 28 501-507 (1984).
2. Schroeder, C., Heider, H., Hegenscheid, G., Schoffel, M., Bubovich, V. I. and Rosenthal, H. A., "An anticholinergic anti-parkinson

- drug norakin selectively inhibits influenza virus replication." *Antiviral Research Suppl I* 95-99 (1985).
3. Ghendon, Y., Markushin, S., Heider, H., Melnikiv, S. and Lotte, V., "Hemagglutinin of influenza A virus is a target for the antiviral effect of norakin." *J Gen Virol* 67 1115-1122 (1986).
 4. Prosch, S., Heider, H., Schroeder, C. and Kruger, D. H., "Mutations in the hemagglutinin gene associated with influenza virus resistance to norakin." *Arch Virol* 102 125-129 (1988).
 5. White, J. M. and Wilson, I. A., "Anti-peptide antibodies detect steps in a protein conformational change: low pH activation of the influenza virus hemagglutinin." *J Cell Bio* 105 2887-2896 (1987).
 6. Doms, R. W., Helenius, A. and White, J., "Membrane fusion activity of the influenza virus hemagglutinin: the low pH-induced conformational change." *JBC* 260 2973-2981 (1985).
 7. Skehel, J. J., Bayley, P. M., Brown, E. B., Martin, S. R., Waterfield, M. D., White, J. M., Wilson, I. A. and Wiley, D. C., "Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion." *PNAS* 79 968-972 (1982).
 8. Sato, Y. B., Kawasaki, K. and Ohnishi, S.-I., "Hemolytic activity of influenza virus hemagglutinin glycoproteins activated in mildly acidic environments." *PNAS* 80 3153-3157 (1983).
 9. Daniels, R. S., Downie, J. C., Hay, A. J., Knossow, M., Skehel, J. J., Wang, M. L. and Wiley, D. C., "Fusion mutants of the influenza virus hemagglutinin glycoprotein." *Cell* 40 431-439 (1985).

10. Hay, A. J., Zambon, M. C., Wolstenholme, A. J., Skehel, J. J. and Smith, M. H., "Molecular basis of resistance of influenza A viruses to amantadine." *J Antimicrob Chemotherapy* 18 Suppl B 19-29 (1986).
11. Ruigrok, R. W. H., Martin, S. R., Wharton, S. A., Skehel, J. J., Bayley, P. M. and Wiley, D. C., "Conformational changes in the hemagglutinin of influenza virus which accompany heat-induced fusion of virus with liposomes." *Virology* 155 484-497 (1986).

Chapter 6

Inhibitor Design Round II

Since the initiation of this project, the DOCK programs have undergone significant modifications [1]. The options for site description have been expanded to allow tailoring of the cluster definition of the site of interest. Between versions 1 and 2 of DOCK, the algorithm has been rewritten to address shortcomings of the earlier code. The revisions enable greater user control over the search, and remedy some of the order dependence inherent in the original program. Scoring has also undergone major renovations. Conversion of atom by atom scoring to a grid-based approach has significantly improved run times of a search. The new routine distinguishes polar and nonpolar atoms in the receptor, and, in DOCK2.1, incorporates electrostatic terms in the scoring function. The ability to evaluate chemical complementarity in addition to shape permits docking of actual compounds instead of molecular templates. The modifications included in the new versions of DOCK are summarized in Chapter 3 and described in detail elsewhere [1].

The other significant advance has been the acquisition of databases of commercially available small molecules. MACCS-II/3D includes three such libraries, the Fine Chemicals Directory (FCD), the Molecular Drug and Data Report (MDDR), and Comprehensive Medicinal Chemistry (CMC). These databases include computer-

generated three-dimensional coordinates for all 75,000 chemicals. The ability to consider commercially available compounds is a significant boon since it sidesteps the difficulties and delays inherent in chemical synthesis.

In addition to code modifications, updated coordinates for several mutant BHAs have become available [2]. However, the adjustment in the positions of the atoms comprising the site is only 1 Å root mean square deviation. This change is not expected to be significant with respect to the resolution of the DOCK method so the original coordinates have been retained in order to compare the Round II results directly with those from Round I.

DOCKing the FCD, MDDR, and CMC Databases

The new DOCK programs and the new databases were used to identify a second round of compounds that might bind to the tail1A site and inhibit the conformational change. First, the shape and electrostatic scoring grids had to be calculated. DISTMAP was used to calculate the scoring grid over the 134 residues that lie within 5 Å of the site. Input parameters used are listed in Table 6-1. Note that while the DISTMAP grid will be used to calculate shape-complementarity scores similar to those computed in DOCK1.1, the scoring functions and thus scores derived from the two programs differ [1, 3]. The electrostatic potential grid was computed by DELPHI over all residues within 40 Å of the tail1A site. A three step focussing method was used with parameters listed in Table 6-2. The two scoring grids were then used in conjunction with DOCK2.1 to dock

the compounds of the FCD, MDDR, and CMC databases to the taillA site. Parameters used in the DOCK run are listed in Table 6-3. Each database was divided into arbitrary sections [4] to be processed in parallel on several computers. Output summaries are listed in Table 6-4.

variable	parameter definition	value
<i>polcon</i>	close contact limit for polar receptor atoms	2.3
<i>ccon</i>	close contact limit for nonpolar receptor atoms	2.6
<i>discut</i>	maximum distance for nonzero score	4.5
<i>maxgrid</i>	grid resolution	3

Table 6-1. Parameter values input to DISTMAP version 1. Run time on an IRIS 4D/35 was 4888.4 seconds.

parameter	value
grid size, step 1	35
grid size, step 2	65
grid size, step 3	65
% box to be filled, step 1	40
% box to be filled, step 2	90
% box to be filled, step 3	150
inner dielectric	2
outer dielectric	80
ionic strength	0.14
debye length	8.143
exclusion radius	1.6
probe radius	1.4

Table 6-2. Input for DELPHI calculation of the electrostatic potential grid surrounding the taillA site.

variable	value
<i>dislim</i>	1.5
<i>nodlim</i>	5
<i>ratiom</i>	0.0
<i>lownod</i>	4
<i>lbinsz</i>	0.15
<i>lovlap</i>	0.0
<i>sbinsz</i>	1.0
<i>sovlap</i>	0.0
<i>nbump</i>	0
<i>natmin</i>	10
<i>natmax</i>	60
<i>expmax</i>	0

Table 6-3. Input parameters for DOCK2.1 used to dock the compounds of the FCD, MDDR, and CMC databases to the tail1A site.

database section	run time (hrs)	total # of ligands	# with positive score	shape scores			best ES score
				min	mean	max	
abc	10:50	3235	2005	12	102	196	-16.3
defg	17:41	5079	3399	7	104	226	-11.6
hijk	NA	5592	4108	14	108	224	-19.5
lmno	46:23	7302	6846	14	120	239	-16.6
pqrs	49:21	6813	6540	17	123	229	-15.4
spns	4:38	1122	895	52	112	206	-5.4
abcd2	45:04	7090	6296	15	120	226	-14.3
efgh2	39:46	7877	6337	17	109	262	-17.2
ijk2	32:55	5580	4842	12	113	206	-16.4
lmn2	45:23	4760	4024	29	115	236	-13.4
spns2	5:38	863	717	49	110	190	-8.1
mddr1	31:43	3272	3228	20	132	246	-14.0
mddr2	33:40	2220	2184	1	129	225	-11.3
mddr3, spns	6:34	3376	3276	21	131	239	-9.5
cmc, spns	38:18	4814	4609	39	127	221	-16.4

Table 6-4. DOCK2.1 output profiles. Sections abc, defg, hijk, lmno, pqrs, spns, abcd2, efgh2, ijk2, lmn2, spns2 contain compounds from

the FCD database; mddr1, mddr2, mddr3, spns comprise the MDDR database; and cmc, spns represents the CMC database. The number of ligands in each section of the database and the time required to DOCK them in cpu hours are listed. Also reported are the number of ligands in each database receiving a positive shape score, the minimum, maximum, and mean shape scores for each run, and the best (most negative) electrostatic (ES) score.

The high scoring ligands were used as indicators of the types of compounds that complement the properties of the site. The top 250-300 molecules from each run from both the shape and electrostatic scoring lists were examined graphically to identify those that represent classes of compounds suitable for biological screening. As in selection of the Round I compounds, particular emphasis was paid to ligands which filled the bulk of the site. Compounds which had good interactions with at least one of the charged groups surrounding the site and/or docked into the histidine pocket were considered especially favorably. Because of the attention to the predominant hydrophobic regions of the site, the shape scores were most discriminatory. Scores in the electrostatic list were low; scores different from zero generally just revealed the presence of one or more charged groups on the molecule, independent of any interactions with residues of the site. However, if the best shape docking coincided with the best electrostatic docking, the compound was given slightly more weight. As with screening of the Round I compounds, ligands with unfavorable solubility or chemical instability were omitted. The selected compounds are listed in Table 6-5.

ID	database section	DOCK list	DOCK rank	shape score	ES score
46	abc	contact	54	153	0.18
38	abc	contact	130	143	-0.92
47	defg	contact	31	171	-3.55
48	defg	contact	67	163	0.99
49	defg	contact	85	158	-1.39
40	defg	elec	24	131	-5.26
37	defg	elec	26	118	-5.06
51	hijk	contact	209	153	-3.48
44	hijk	contact	2	203	-0.87
39	hijk	contact	37	176	0.31
43	hijk	elec	9	126	-9.30
42	hijk	elec	23	128	-6.76
58	lmno	elec	52	151	-5.95
35	pqrs	contact	123	173	1.89
57	pqrs	elec	41	128	-4.97
55	fcd.spns	contact	4	184	-1.07
54	abcd2	contact	221	166	0.20
56	fcd2.spns	contact	33	152	-1.70
50	efgh2	contact	105	171	-1.00
41	efgh2	elec	92	139	-4.43
53	lmn2	contact	297	156	1.15
45	lmn2	elec	9	117	-9.34
52	ijk2	contact	11	191	-2.06

Table 6-5. Compounds from the FCD selected for biological screening. Structures are depicted in Table 6-11. The section of the FCD database from which each compound is derived (see Table 6-4) is listed. The orientation of each ligand receiving the highest shape score and that receiving the best electrostatic score were saved as independent lists. "DOCK list" indicates whether the ligand was selected based on its best contact scoring orientation or its best electrostatic scoring orientation. The "DOCK rank" is how well that orientation scored compared to the other ligands in the same database section, on the same list. Compounds ranked 1 received the best score. The electrostatic (ES) score of the compounds selected from the contact score list were calculated for the orientation receiving the best shape score, and vice versa.

DOCKing excerpted databases

The results of the initial FCD docking were used to select classes of compounds with good complementarity to the site to dock in more detailed runs. A higher resolution definition of the site was achieved by running SPHGEN on all surface points within 6 Å of the site and saving all the calculated spheres instead of just one per atom. To focus on the central region of the site, the spheres with centers farther than 5 Å of the center of mass of the site were eliminated. CLUSTER, the program that allows tailoring of the sphere sets, was then run on the remaining spheres to give a cluster composed of 27 spheres. Parameters input to CLUSTER are listed in Table 6-6.

variable	value
<i>maxrad</i>	5.0
<i>m2xrad</i>	0
<i>povlap</i>	0
<i>clusiz</i>	250
<i>minsiz</i>	1
<i>minflg</i>	0
<i>yn</i>	y
<i>minrad</i>	1.0
<i>rincr</i>	0.5
<i>nearnb</i>	0.5
<i>nearad</i>	0.25

Table 6-6. Parameters for the CLUSTER runs.

A new scoring grid was calculated by DISTMAP2 using the parameters in Table 6-7. The resolution was increased from 3 grid points/Å² to 4. Also, the close contact limits were increased to be more representative of the distances observed in the crystallographic complex of an inhibitor bound to a hydrophobic pocket [5].

variable	parameter definition	value
<i>polcon</i>	close contact limit for polar receptor atoms	2.4
<i>ccon</i>	close contact limit for nonpolar receptor atoms	2.8
<i>discut</i>	maximum distance for nonzero score	4.5
<i>maxgrid</i>	grid resolution	4

Table 6-7. Parameter values input to DISTMAP version 2. Run time on an IRIS 4D/35 was 142 seconds.

Instead of redocking the entire FCD, the results of the full run were used to select a subset of the chemicals. Excerpted databases contained classes of compounds with shape and electrostatic complementarity to the central part of the site, chemical properties dictated by the screening assays, and for which a variety of derivatives are either commercially available or easily synthesized. These compounds were collected into short databases with the substructural searching facility of the MACCS program. The database "ntsearch" contains ligands with a central 5-membered ring that is not completely aromatic. "Cyclo" ligands contain a central 5- or 6-membered ring that is part of a more complicated ring system. Compounds in the "4chgs" database contain four polar atoms at the

appropriate spacings to interact with the charged groups surrounding the site. They require a nitrogen-nitrogen distance of 2.9 - 9.5 Å, a nitrogen and oxygen separated by up to 4.5 Å, and a second nitrogen - oxygen distance within 2.4 - 10 Å. A fourth polar atom was required to be 3 - 8 Å from the plane defined by the other three atoms. The parameters used to DOCK these databases to the higher resolution definition of the site are listed in Table 6-8, and the output is summarized in Table 6-9. Compounds selected for biological screening using the criteria described above and in Chapter 4 are given in Table 6-10.

variable	value
<i>dislim</i>	1.5
<i>nodlim</i>	5
<i>ratiom</i>	0.0
<i>lownod</i>	4
<i>lbinsz</i>	0.15
<i>lovlap</i>	0.0
<i>sbinsz</i>	1.0
<i>sovlap</i>	0.0
<i>nbump</i>	0
<i>natmin</i>	10
<i>natmax</i>	35
<i>nsav</i>	500
<i>nnsav</i>	200
<i>expmax</i>	1
<i>fctbmp</i>	1

Table 6-8. Input parameters used to dock the excerpted FCD databases using DOCK2.1.

database	# of comps	# with positive score	time (min)	contact score			best ES score
				min	mean	max	
ntsearch	422	337	3:05	61	134	175	-12.5
cyclo	437	320	3:48	65	135	187	-13.8
4chgs	4144	3171	25:03	51	126	201	-30.7

Table 6-9. Output profiles from DOCKing selected compounds from the FCD. Classes of compounds contained in each database are described in the text. The cpu time for running on an IRIS 4D/35 is listed. The number of ligands docking with positive contact scores, and the minimum, mean, and maximum shape scores for each run are reported. Also listed is the best electrostatic (ES) score received by any compound in the database.

ID	database	DOCK rank	shape score	ES rank	ES score
75	ntsearch	4	166		
79	ntsearch	11	161		
67	ntsearch	14	161		
76	ntsearch	18	159		
77	ntsearch	20	158		
62	ntsearch	30	156		
83	ntsearch	47	152		
59	ntsearch	54	151		
82	ntsearch	55	151		
66	ntsearch	71	148		
74	ntsearch	82	147		
72	ntsearch	83	145		
70	ntsearch	97	143		
78	ntsearch	98	143		
65	ntsearch	112	141		
80	ntsearch	113	141		
73	ntsearch	119	140		
63	ntsearch	120	140		
69	ntsearch	122	139		
60	ntsearch	128	139	1	-12.5
61	ntsearch	129	139		
81	ntsearch			4	-3.8
71	cyclo	1	187		
64	cyclo	18	163		
68	cyclo	55	149		

Table 6-10. Compounds from the ntsearch, cyclo, and 4chgs databases which were selected for biological screening. Structures of each molecule are depicted in Table 6-11. Ligands selected on the basis of the orientation receiving the best shape score are listed with their rank on that list and the contact score received. Ligands selected by their electrostatically ranked orientations (ES) are listed with their ES rank and score.

Table 6-11. Round II compounds.

ID	name	source	FW	structure
35	citric acid trimethyl ester	American Tokyo Kasei	234	
36	4,4'-isopropylidene bis(2,6-dibromophenol)	Aldrich	544	
	α -methyl-DL-glutamic acid	Sigma	161	
38	4,4'-isopropylidene diphenol	Aldrich	228	
39	N,N-diphenyl benzamidine	Lancaster Synthesis	176	
40	m-tolylacetic acid	Aldrich	150	
41	3-methyl-1-adamantane acetic acid	Aldrich	208	
42	3-(diethylamino)propionic acid hydrochloride	Aldrich	182	
43	[1R-(2-endo,3-exo)]-3-hydroxy-4,7,7-trimethyl-bicyclo[2.2.1]-heptane-2-acetic acid	Aldrich	212	

Table 6-11, continued.

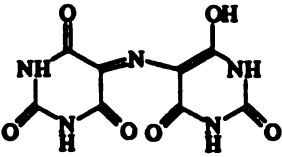
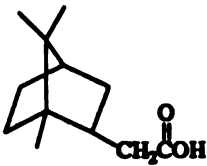
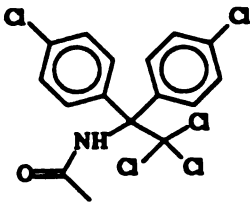
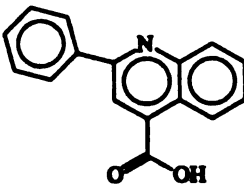
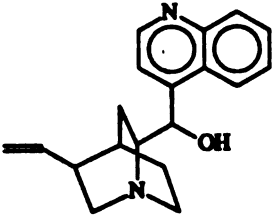
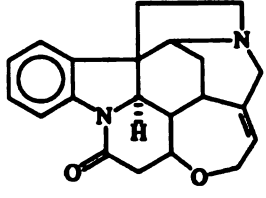
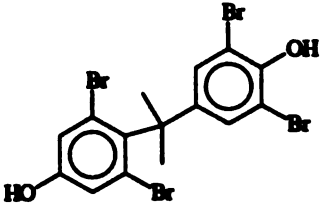
44	murexide	Sigma	284	
45	(-) isoborneol-3-acetic acid	Sigma	212	
46	N-[1,1-bis(p-chlorophenyl)-2,2,2-trichloroethyl]acetamide	Aldrich	411	
47	2-phenyl-4-quinoline carboxylic acid	Aldrich	249	
48	cinchonine	Sigma	294	
49	strychnine	Sigma	334	
50	tetrabromophenol	Sigma	544	

Table 6-11, continued.

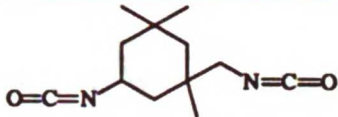
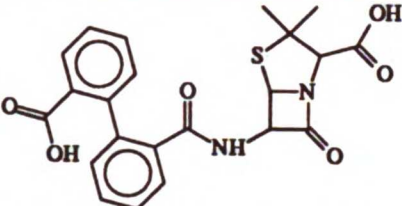
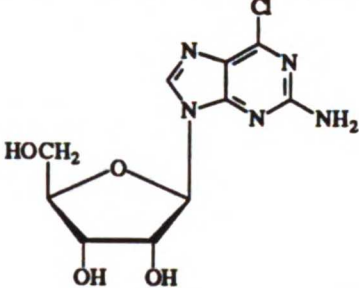
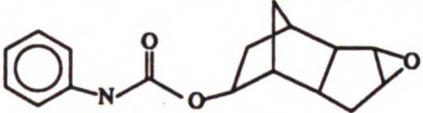
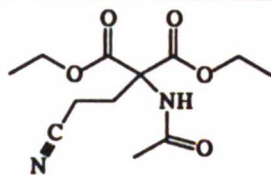
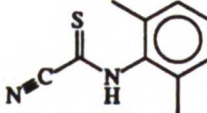
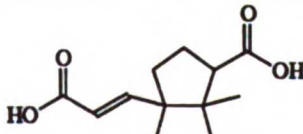
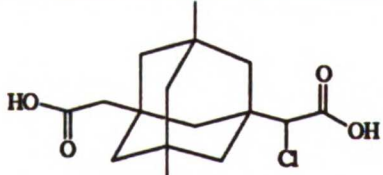
51	isophorene diisocyanate	Sigma	222	
52	2-(2'-carboxyphenyl)benzoyl-6-aminopenicillanic acid Na ⁺	Sigma	484	
53	2-amino-6-chloropurine riboside	Sigma	302	
54	1A,1B,3,4,5,5A,6,6A-octahydro-25-methano-2-hindeno-(12B)-oxiren-3-yl-N-(ptolyl) carbamate	Bader	299	
55	diethyl 2-acetamido-2-(2-cyanoethyl)-malonate	Bader	270	
56	N-(2,6-xylyl) cyanothioformamide	Bader	190	
57	3-carboxy-1,2,2-trimethylcyclopentane acrylic acid	Bader	226	
58	α -3-chloro-5,7-dimethyl-1,3-adamantane diacetic acid	Bader	315	

Table 6-11, continued.

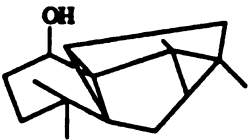

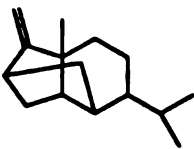
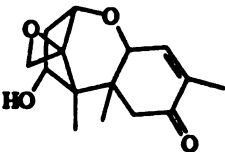
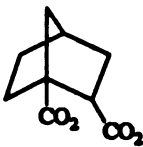
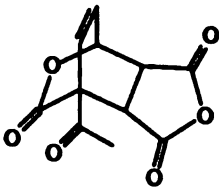
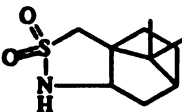
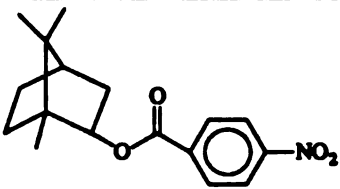
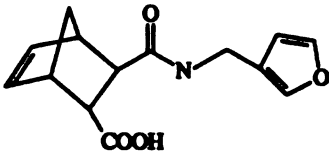
59	(+)-cycloisolongifol-5-ol	Fluka	220	
60	(-)-isolongifolene	Fluka	204	
61	(-)-isosativene	Fluka	204	
62	trichothecolone	Sigma	264	
63	bicyclo (2.2.1) heptane-3,4-dicarboxylic acid	Bader	184	
64	7-acetyl-4,9-dioxatricyclo (5,2,2,0 ^{2,6}) undec-10-ene-3,5,8-triene	Bader	236	
65	L-(+)-2,10-camphorsultam	Aldrich	215	
66	isobornyl-4-nitrobenzoate	Bader	303	
67	3-(N-(3-furfuryl)carbamoyl-5-norbornene-2-carboxylic acid	Bader	261	

Table 6-11, continued.


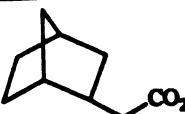
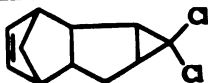
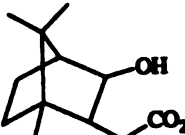
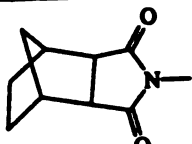

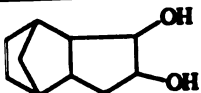
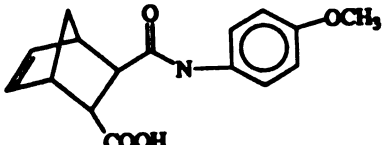
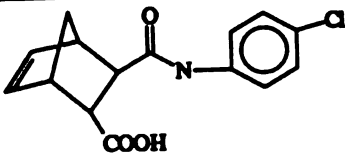
68	pentacyclo[5.4.0.0.0.0]-undecane-8,11-dione	Aldrich	174	
69	2-norbornane acetic acid	Aldrich	154	
70	5,5-dichlorotetracyclo(6.2.1.0.2,7.0.4.6) undec-9-ane	Bader	215	
71	1R-2-endo-3-exo-3-hydroxy-4,7,7-trimethyl-bicyclo[2.2.1]heptane-2-acetic acid	Aldrich	212	
72	N-methyl-5-norbornene-2,3-dicarboximide	Bader	177	
73	2-chloro-7,7-dimethyl-1-norbornane ethanol	Bader	203	
74	tricyclo(5.2.1.0.2,6)decane-3,4-diol	Bader	168	
75	3-(N-(4-methoxyphenyl) carbamoyl)-5-norbornene-2-carboxylic acid	Bader	287	
76	3-(N-(4-chlorophenyl) carbamoyl)-5-norbornene-2-carboxylic acid	Bader	292	

Table 6-11, continued.

77	2-(N-(4-methylphenyl) carbamoyl-5-norbornene-2-carboxylic acid)	Bader	271	
78	1,2,3,4,4A,5,8,8A-octahydro-1,4:5,8-dimethanonaphth-2-yl acetate	Bader	218	
79	4,8,-bis(hydroxymethyl) tricyclo[5.2.1.0]decane	Aldrich	196	
80	1,4,4A,4B,5,8,8A,8B,9,10-decahydro-1,4:5,8-dimethanoanthracene-9,10-dione	Bader	240	
81	anti-3-oxotricyclo[2.2.1.0] heptane-7-carboxylic acid	Aldrich	152	
82	bornyl isovalerate	Pfaltz & Bauer	238	
83	4A,5,8,8A-tetrahydro-5,8-methano-1,4-naphthoquinone	Pfaltz & Bauer	174	

Table 6-11 notes. Assigned ID number, name, commercial supplier, formula weight, and chemical structure are listed for each.

Summary

DOCK2 was used to search a database of commercially available compounds for ligands likely to bind to the tail1A site on hemagglutinin. Selected compounds had steric and electrostatic properties complementary to those of the target site.

References

1. Shoichet, B. K., Bodian, D. L. and Kuntz, I. D., "Molecular docking using shape descriptors." *J Comp Chem* 13 380-397 (1992).
2. Weis, W. I., Brunger, A. T., Skehel, J. J. and Wiley, D. C., "Refinement of the influenza virus hemagglutinin by simulated annealing." *JMB* 212 737-761 (1990).
3. DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R., "Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure." *J Med Chem* 31 722-729 (1988).
4. Meng, E. C., personal communication.
5. Badger, J., Minor, I., Oliveira, M., Smith, T. and Rossmann, M., "Structural analysis of antiviral agents that interact with the capsid of human rhinoviruses." *Proteins* 9 1-19 (1989).

Chapter 7

Experimental Testing of Round II Compounds

The strategy used for screening compounds from the second round of modeling is somewhat different from that of Round I. Software modifications and improvements in experimental techniques eliminated the two most time-consuming and difficult steps of the earlier approach. Searching databases of commercially available compounds minimized dependence on potentially long and troublesome chemical syntheses. Selection of the scintillation proximity assay as the primary screening assay increased both the reliability and convenience of the screen. These combined developments allowed the relatively easy acquisition and testing of the Round II compounds.

Screening Round II compounds

The primary assay used for screening the Round II compounds was the scintillation proximity assay (SPA). Both immunoprecipitation (IP) and SPA recognize the low pH form of hemagglutinin by its ability to bind conformation-specific antibodies. The assays differ in how they detect the BHA-antibody complexes.

Figure 7-1 compares the precipitation¹ of low pH BHA with the α -fusion peptide antibody by the two techniques. In a mixture of BHA in various conformations the α -fusion peptide antibody will specifically bind to hemagglutinin with exposed fusion peptides. In IPs, the antibody and any associated protein is then bound to protein A-agarose beads. Native BHA, unlinked to the beads, is washed away with a high salt buffer. All radioactivity remaining after the washes is counted in scintillation fluid. In contrast, differentiation of neutral and low pH forms of BHA by SPA does not depend on their physical separation. The experiment exploits the distance-dependent quenching of low energy β -emitters by aqueous media. As in immunoprecipitation, the α -fusion peptide antibody is bound to both radiolabeled low pH BHA and to protein A. However, in the SPA protein A is coupled to microspheres containing scintillant. Only BHA bound to antibody-bead complexes will trigger emission of light by the fluorophore. Energy produced by unbound protein will be dissipated in the solvent and thus remain undetected.

Since it does not require wash steps to remove unbound protein, the SPA is both easier and more reproducible than the IP assay. In addition, the harsh washing conditions of the IP inevitably dissociate some BHA-antibody-protein A complexes and reduce the efficiency of the precipitation. This is particularly true for low affinity antibodies such as the α -fusion peptide antibody. With the washes omitted, counting in the SPA is done under equilibrium

¹Detection of BHA-antibody complexes by either IP or SPA will be referred to as precipitation even though the complexes are only centrifugally pelleted in the IP assay.

conditions so much smaller amounts of antibody are required to achieve similar levels of precipitation.

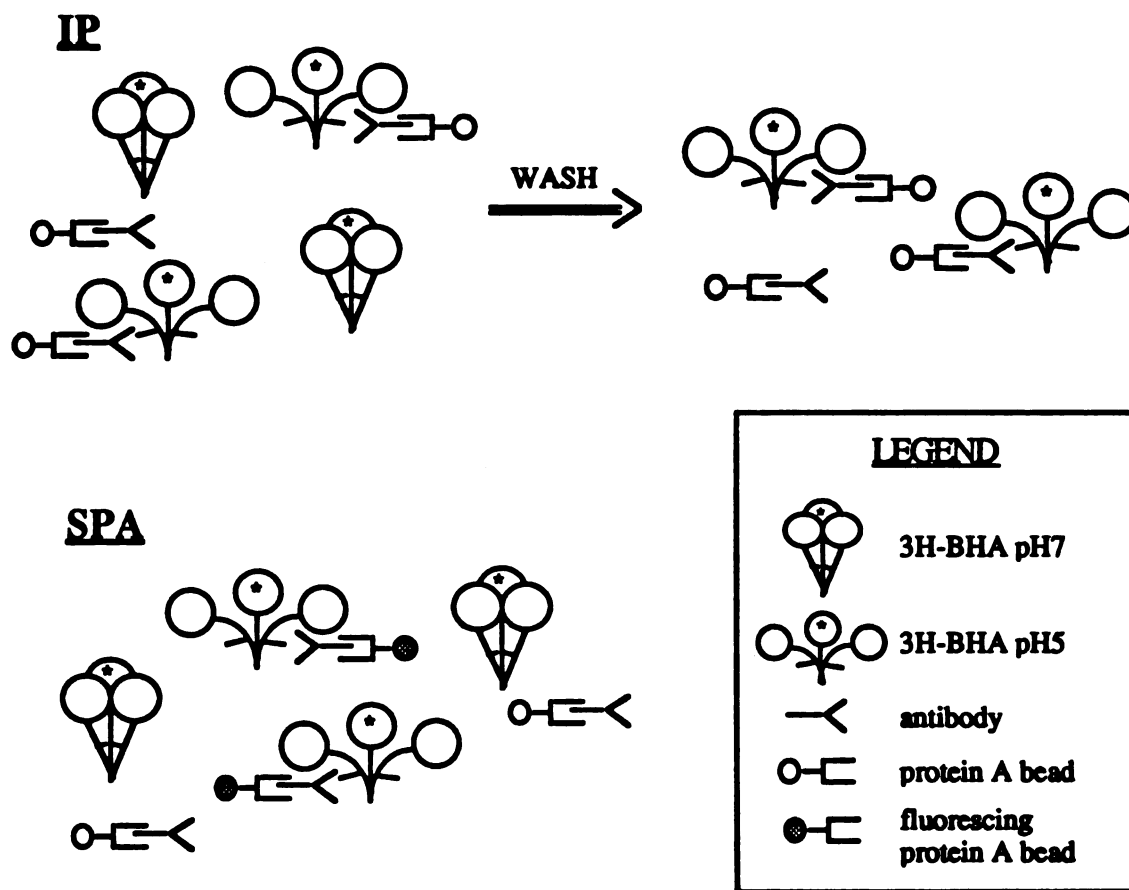


Figure 7-1. Comparison of the immunoprecipitation (IP) and scintillation proximity (SPA) assays. Low pH trimers are distinguished from neutral pH trimers by their ability to bind the α -fusion peptide antibody. In the IP, primary antibody and its associated BHA bind to protein A-agarose beads. Unbound protein is washed away and the remaining radioactivity assessed by scintillation counting. In the SPA, primary antibody and its associated BHA bind to fluorophore-coupled protein A beads. Only bound BHA will trigger fluorescence of the scintillant.

However, since they are not washed away, the compounds being screened by SPA are present during scintillation counting. Some of the chemicals, particularly those that are highly colored, quench the detection of radioactivity. Two different controls distinguish whether apparent inhibition arises from inhibition of the conformational change or from nonspecific effects. In the first control, the samples are collected and washed under mild conditions. An increase in cpm following the rinse is indicative of nonspecific quenching. However, this method does not work well for compounds with low solubility in aqueous media and does not produce reliable and quantitative results. In the second control ("POST"), the test compound is added to a sample of BHA that had previously been acidified and reneutralized. The mixture is then precipitated as usual. Since the BHA is known to be in the low pH conformation, a reduction in measurable cpm compared to controls with ligand omitted proves the compound interferes with the detection system.

The commercial compounds selected from the FCD were screened by SPA using the α -fusion peptide antibody. The results are shown in Table 7-1. POST controls were carried out for compounds which reduced the cpm precipitated by at least 10%. Highly toxic compounds or compounds insoluble in DMSO or water were not tested.

compound	% inhibition	sample %sd	POST
35	-5.9	0.1	nd
38	-6.6	0.3	nd
39	4.3	0.8	nd
40	1.2	0.2	nd
41	1.4	0.2	nd
42	-1.0	0.3	nd
43	0.3	3.0	nd
44	21.3	2.6	22.2
45	-0.8	1.7	nd
47	22.7 ± 10.6	4.9	17.9
49	-0.5	3.3	nd
52	15.2 ± 10.1	4.8	6.3
53	-3.1	0.1	nd
54	0.3	1.7	nd
55	-1.6	1.9	nd
56	12.7	3.0	nd
57	0.1	0.1	nd
58	-3.7	2.3	nd
60	-3.3	2.8	nd
61	0.1	6.5	nd
63	-3.9	0.7	nd
64	1.1	3.0	nd
65	1.0	3.3	nd
67	4.0	5.6	nd
68	1.1	2.2	nd
69	-13.8	7.8	nd
70	7.6	3.7	nd
71	0.4	4.4	nd
72	1.1	0.6	nd
73	8.1 ± 3.2	4.5	4.0
74	0.4	2.8	nd
75	0.0	1.0	nd
76	17.7 ± 1.3	1.9	13.2
77	10.2 ± 0.4	2.1	4.9
78	5.6	0.1	nd
79	9.7	2.5	nd
81	-1.8	3.8	nd
82	15.4 ± 15.1	4.6	5.0
83	49.1 ± 15.0	2.1	4.2

Table 7-1. Effect of Round II compounds in the SPA with the α -fusion peptide antibody. All compounds were tested in duplicate at 1 mM. “% inhibition” is the % reduction in cpm precipitated compared to control samples with test ligands omitted. The “sample %sd” is the standard deviation in % inhibition for duplicate reactions

within a single experiment. Data for compound 47 is averaged from 3 independent experiments; 52, 2; 73, 2; 76, 2; 77, 2; 82, 3; 83, 16. "POST" represents the % reduction in cpm observed in compound-containing samples of low pH BHA.

Several compounds reduced the detectable cpm by at least 15%. The apparent inhibition by compounds 44, 47 and 76 is accounted for by the POST controls. The POST values for compounds 52 and 82 also lie within the range of inhibition observed. In contrast, the activity of compound 83 could not be attributed to nonspecific quenching. To study the effect of compound 83 in more detail, the activity as a function of concentration was determined (Figure 7-2). Half-maximal inhibition fell between 10^{-3} M (174 $\mu\text{g/ml}$) and 10^{-4} M (17.4 $\mu\text{g/ml}$).

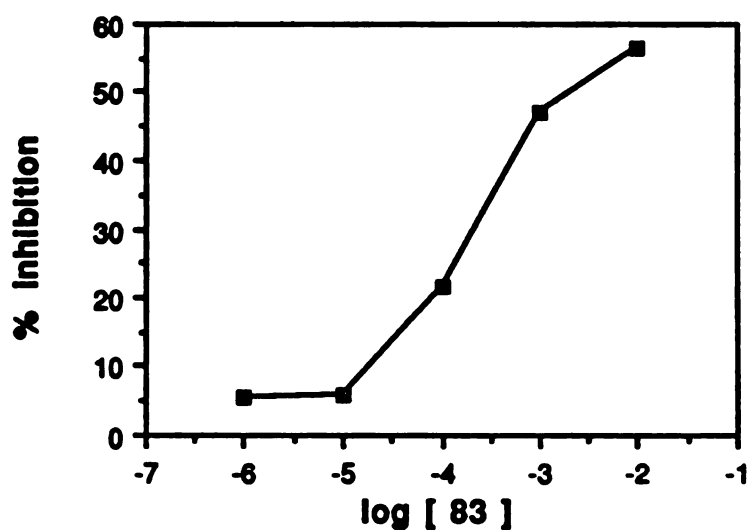


Figure 7-2. Dose response of compound 83 in the SPA. % inhibition is averaged over duplicate precipitations from 1 - 3 independent experiments. The mM point is the average of 16 independent determinations (Table 7-1). Values are corrected for nonspecific

quenching as described in Chapter 3. Inhibition of the POST controls was 26.4% at 10 mM and 4% or less at all other concentrations. Standard deviations are: 1 μ M, 9.5; 10 μ M, 19.6; 0.1 mM, 17.1; 1 mM, 15.0. These errors reflect the daily variability of the activity at each concentration.

Since compound 83 at concentrations below 10 mM does not reduce precipitation in POST controls, the observed inhibition is presumably due to an interaction with hemagglutinin. To distinguish whether the effect is a result of noncovalent binding or a covalent modification, a dilution experiment was carried out. If the activity is due to noncovalent binding it should be reversible by dilution; covalent modifications are less likely to be reversed under these conditions. Compound 83 was incubated with BHA at an initial concentration of 10^{-3} M. Half the mixture was then diluted to 10^{-5} M, a concentration at which negligible inhibition had been observed (Figure 7-2). The other half was diluted to the same volume, but the concentration of compound 83 was maintained at 10^{-3} M. Acidification and precipitation reactions were carried out as usual. The results in Table 7-2 reveal that the inhibition caused by compound 83 at 1 mM is reversible by dilution, suggesting that the interaction with BHA is noncovalent.

sample	conc (M)	% inhibition
1	$10^{-3} \rightarrow 10^{-3}$	45.3
2	$10^{-3} \rightarrow 10^{-5}$	4.6

Table 7-2. SPA dilution experiment with compound 83. Samples were preincubated with 1 mM 83. Sample 1 was maintained at 10^{-3}

M throughout the experiment but sample 2 was diluted to 10^{-5} M prior to acidification. All samples were incubated at low pH for 5 minutes then precipitated by SPA with the α -fusion peptide antibody. "% inhibition" is the % reduction in cpm precipitated compared to controls with 83 omitted. POST controls showed no nonspecific quenching by 83 at these concentrations.

However, the dilution experiment does not rule out the possibility that compound 83 has an irreversible effect caused by acidification. To address this, inhibition as a function of time at low pH was determined. If the effect is due to chemical reactivity at pH 5, inhibition should remain constant or increase during extended incubations. If the effect of compound 83 is noncovalent in nature then, due to the irreversibility of the conformational change itself, the inhibition would be expected to decrease as a function of time. The results for 10 minutes of exposure to pH 5.0 are shown in Figure 7-3. Initially, at 10 seconds, 73% inhibition is observed. After 10 minutes, inhibition has decreased to 43%. This implies that inhibition is reversible at low pH. Combined with the results of the dilution experiment, the data suggest that compound 83 is a reversible inhibitor.

There are interesting biological implications of this experiment. In order for our antiviral strategy to be successful, the inhibitors must prevent fusion in the acidic endocytic compartments until the virus is inactivated in the lysosome². For the WSN strain of

²The antiviral compound does not necessarily have to prevent conformational changes along the entire endocytic pathway. It is possible that hemagglutinin reaching a pre-lysosomal compartment with a pH below the fusion threshold will undergo rearrangements that do not trigger membrane fusion.

influenza, the mean time from endocytosis to degradation of the hemagglutinin is approximately 20 minutes longer than the average time (~25 minutes) for endocytosed virions to encounter the pH threshold of fusion [1]. As shown in Figure 7-3 for C22 BHA, compound 83 still maintains 60% of its maximal inhibition of fusion peptide exposure even after 10 minutes of exposure to pH 5.0. This suggests that 83 or a stronger analog may be able to prevent fusion long enough to inhibit viral infection.

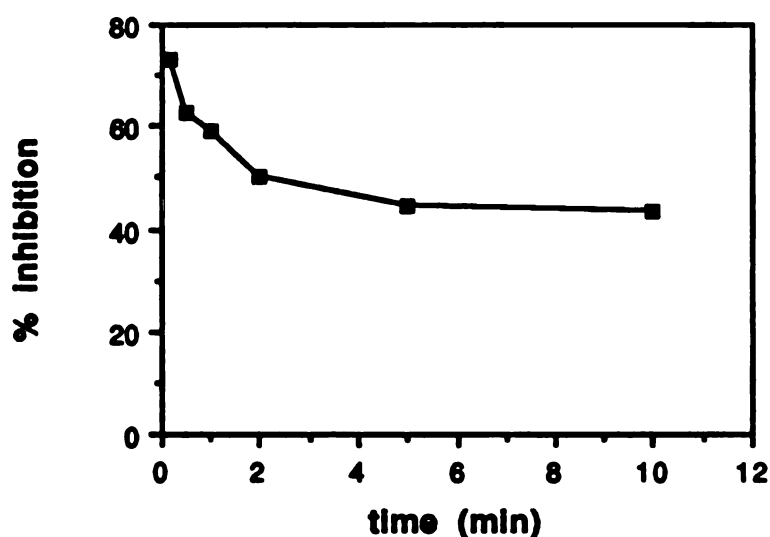


Figure 7-3. Inhibition by compound 83 as a function of time at low pH. The cpm precipitable by SPA with the α -fusion peptide antibody in the presence of 1 mM 83 are compared to those in samples without 83. Background counts have been subtracted from each value. POST samples gave similar results to the samples with 83 omitted.

Testing Derivatives of the Lead Compound

Scintillation Proximity Assay

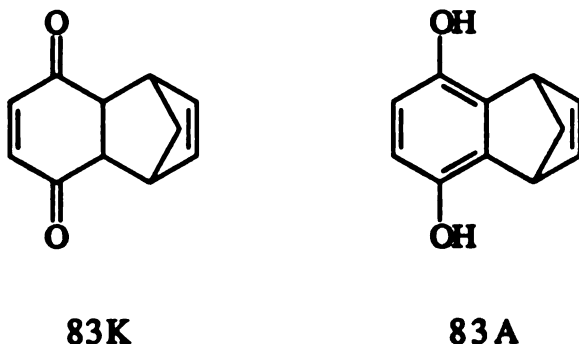


Figure 7-4. Structures of compounds 83K and 83A.

The chemical structure of 83 reveals the potential of the compound to tautomerize to an alternate form. The structures of the two isomers, 83K and 83A, are shown in Figure 7-4. The commercially available 83 may be a mixture of the two forms. In order to determine which is the active species, the two compounds were prepared separately (Appendix A) and tested individually. The precipitation of BHA in the presence of each of the two purified components as well as with the original commercial sample are compared in Figure 7-5. Inhibition was observed with all three species. The concentration profile of 83K exactly mimics that of 83, but compound 83A is active at 100-fold lower concentrations. Although this suggests that both compounds are active, it does not rule out the possibility that the activity observed with 83 and 83K is due to contamination by 83A, possibly generated during the course of the experiment.

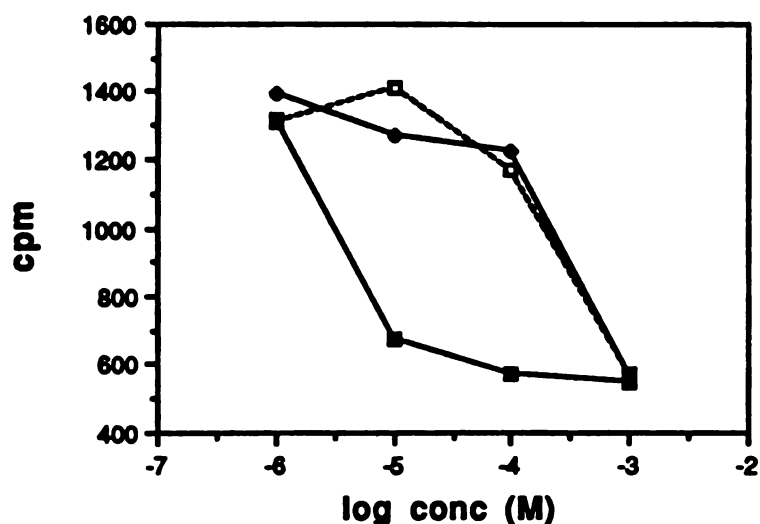


Figure 7-5. Effect of compounds 83 (\square), 83K (\diamond), and 83A (\blacksquare) on precipitation of BHA by SPA with the α -fusion peptide antibody. Reactions were carried out as described in the methods. Cpm precipitated are averaged from duplicate samples from a single experiment.

The fortuitous increase in inhibitory activity observed with 83A encouraged us to test other analogs of 83. Compounds were found by substructure searching using MACCS-II/3D. Chemicals selected represent the commercially available variety of substituents. The structures of these compounds are shown in Table 7-7.

The derivatives were tested for their ability to inhibit fusion peptide exposure, as measured by SPA with the α -fusion peptide antibody. They span a range of activities, from inactive at mM concentrations to active in the μ M range (Table 7-3). Controls showed that none of the compounds tested (84, 89, 90, 91, 99, 116,

117, 118, 119, 121, and 135 at 1 mM and 114 and 124 at 0.1 mM) affect the pH of the reactions. The approximate concentrations at which half-maximal inhibition was observed (IC₅₀) for each ligand are listed in Table 7-6.

Table 7-3. Derivatives of 83 in the SPA.

compound	pure	log conc (M)	sample		POST control	
			% inhibition	n	% inhibition	n
83K	*	-3	59.5 ± 3.7	3	-5.3 ± 2.1	2
		-4	6.6 ± 4.4	3	3.0	1
		-5	0.9 ± 4.2	3	0.8 ± 12.7	2
		-6	-2.5 ± 2.1	3	-6.2	1
83A	*	-3	54.3 ± 17.7	6	0.5 ± 1.1	2
		-4	48.6 ± 19.7	6	-0.9 ± 1.5	2
		-5	18.7 ± 21.1	5	-0.5	1
		-6	6.1 ± 9.1	4	10.8	1
84		-3	41.2 ± 13.6	2	15.6	1
		-4	8.8	1	nd	0
		-5	-5.2	1	nd	0
		-6	-2.8	1	nd	0
85		-3	1.9	1	nd	0
86		-3	-20.2	1	nd	0
87		-3	6.3	1	nd	0
88		-3	17.9	1	nd	0
89		-3	49.9	1	53.4	1
		-4	19.4	1	nd	0
90	*	-3	86.4	1	68.2	1
		-4	47.7	1	30.8	1
		-5	4.3	1	4.6	1
		-6	-3.8	1	nd	0
91		-3	88.6	1	54.5	1
		-4	66.5	1	nd	0
		-5	61.7	1	nd	0
		-6	15.0	1	nd	0
92		-3	7.6	1	nd	0
94		-3	-4.8	1	nd	0
95		-3	-2.5	1	nd	0
96		-3	6.1	1	nd	0
97		-3	71.6 ± 8.7	4	59.5	1
		-4	74.7 ± 26.3	2	nd	0
		-5	37.9 ± 14.9	2	nd	0
		-6	3.1 ± 0.9	2	nd	0
98		-3	-3.3	1	nd	0

Table 7-3, continued.

99	*	-3	81.7	1	79.7	1
		-4	65.6	1	66.5	1
		-5	24.9	1	nd	0
		-6	2.2	1	nd	0
100		-3	6.4	1	nd	0
101		-3	11.3 ± 9.1	2	nd	0
		-4	2.0 ± 4.0	2	nd	0
		-5	-2.1 ± 2.8	2	nd	0
102		-3	10.5 ± 9.8	2	nd	0
103		-3	23.7	1	15.7	1
104		-3	24.6 ± 13.7	2	nd	0
		-4	11.2	1	nd	0
105		-3	48.1 ± 26.0	2	nd	0
		-4	26.8	1	nd	0
		-5	24.4	1	nd	0
		-6	3.6	1	nd	0
106		-3	11.3	1	nd	0
107		-3	49.5 ± 22.7	3	nd	0
		-4	25.6	1	nd	0
		-5	-9.7	1	nd	0
		-6	-1.9	1	nd	0
108		-3	-0.0	1	nd	0
109		-3	0.1	1	nd	0
110		-3	75.7 ± 9.1	4	nd	0
		-4	32.7 ± 2.3	2	nd	0
		-5	12.9 ± 4.0	2	nd	0
		-6	-9.4 ± 5.6	2	nd	0
111	*	-3	65.7 ± 1.6	2	13.3	1
		-4	70.5	1	9.5	1
		-5	26.6	1	nd	0
		-6	8.0	1	nd	0
112		-3	40.1	1	nd	0
114	*	-5	50.4	1	33.1	1
		-6	23.5	1	5.6	1
		-7	-14.6	1	1.9	1
115		-3	62.1 ± 5.4	2	nd	0
		-4	15.4	1	nd	0
		-5	-19.6	1	nd	0
		-6	-11.0	1	nd	0
116	*	-3	82.2	1	82.7	1
		-4	64.7	1	69.2	1
		-5	28.5	1	nd	0
		-6	2.0	1	nd	0
117	*	-3	79.3 ± 15.4	3	-3.4	1
		-4	80.3 ± 1.5	2	2.9	1
		-5	46.4 ± 21.8	2	9.1	1
		-6	-0.1 ± 18.7	2	4.7	1

Table 7-3, continued.

118	*	-3	64.3 ± 14.4	3	61.4	1
		-4	37.7 ± 6.7	2	50.5	1
		-5	22.4 ± 19.0	2	15.3	1
		-6	4.0	1	nd	0
119	*	-3	90.7	1	95.8	1
		-4	54.5	1	42.9	1
		-5	6.2	1	9.9	1
		-6	-5.7	1	2.4	1
120	*	-3	57.9	1	54.0	1
		-4	7.8	1	7.7	1
		-5	1.3	1	0.2	1
121	*	-3	81.6 ± 8.2	3	83.6	1
		-4	61.7 ± 15.2	2	77.5	1
		-5	17.9 ± 6.1	2	nd	0
		-6	-35.2	1	nd	0
122		-3	15.7	1	nd	0
124		-3	86.9 ± 5.4	2	nd	0
		-4	79.8 ± 9.6	2	87.0	1
		-5	62.9 ± 4.8	2	72.8	1
		-6	31.9	1	37.8	1
125		-3	64.8	1	nd	0
		-4	53.0	1	nd	0
		-5	1.8	1	nd	0
		-6	-20.9	1	nd	0
126	*	-3	36.1	1	3.2	1
		-4	21.1	1	2.2	1
		-5	-1.4	1	nd	0
127		-3	-9.7	1	nd	0
135	*	-3	64.0	1	56.2	1
		-4	27.9	1	29.5	1
		-5	-1.6	1	nd	0

Table 7-3 notes. Activity of derivatives of compound 83 in the SPA with the α -fusion peptide antibody. “% inhibition” is the % reduction in cpm compared to controls without trial ligand. All precipitations were done in duplicate. For samples repeated in independent experiments the mean % inhibition, its standard deviation, and the number of experiments averaged (n) are reported. Results from POST controls were treated similarly. “nd” not done. Stock solutions of all compounds were prepared in DMSO. Compounds 93, 113 and 123 were insoluble in DMSO and were not tested. Compounds 101, 102, 105, 106, 107, and 110 were only partially soluble at 1 mM. “*” indicates compounds known to be pure (Appendix A). Impure compounds 91, 97 and 125 and unstable 110 were not pursued. Data could not be obtained for higher concentrations of 114 due to complete quenching of the samples.

Averaged over 5 independent experiments, the maximum inhibition for nonquenching compounds was $65.6 \pm 7.8\%$. No compound was able to reduce the cpm precipitated to background levels. The inhibitory activity could not be increased by raising the concentration of ligand (see for example 83A, 111, and 117 in Table 7-3), and the maximum observed inhibition was similar for sufficiently potent compounds. Since the BHA was >90% pure and no counts were precipitated without prior acidification, the residual 34% is not due to precipitation of a radiolabeled contaminant. There are several possible explanations for the lack of complete inhibition. The compounds may only be 66% effective at inhibiting the conformational change due to weak interactions with the protein. The compounds may be binding to a site unable to regulate fusion peptide exposure completely. If 100% inhibition requires interactions with 2 or more sites per trimer, the compounds may not be binding to a sufficient number of sites concurrently. Alternatively, a fraction of the BHA trimers may be misfolded or otherwise uninhibitable.

Hemagglutination

Compounds inhibiting the conformational change of hemagglutinin were tested for their ability to inhibit fusion and infectivity. Assays measuring these processes depend on viral recognition of cellular receptors. To rule out the possibility that any observed inhibition in these experiments is due to inhibition of

virus-cell binding, hemagglutination tests were performed with some of the compounds. Derivatives 83 (1 mM), 91 (1 mM), 97 (0.1 mM), 110 (1 mM), and 117 (10 μ M) clearly had no effect on binding between intact virions and their receptors on red blood cell membranes.

Hemolysis

The results from the SPA experiments suggest that compound 83 and several derivatives can inhibit exposure of the fusion peptide. If so, they should also inhibit fusion. The hemolysis assay has been used previously to study the membrane fusion activity of hemagglutinin (see Chapter 5). Influenza viruses rupture the membranes of red blood cells under fusion-inducing conditions. The amount of hemoglobin released, as measured by the absorbance at 570 nm, is proportional to the amount of fusion. Therefore, compounds capable of preventing hemagglutinin-mediated fusion should reduce the OD by an amount related to the inhibitory activity.

There are two potential problems with using the hemolysis assay to screen compounds. 1) Some compounds can lyse rbc's, even in the absence of virus. Control samples in which rbc's were incubated with compounds under the conditions of the hemolysis experiment revealed the extent of compound-induced lysis. The OD of these control samples was then subtracted from the OD of samples containing virus plus compound. 2) The compounds may alter the absorption spectrum of hemoglobin. The ability of the compounds to

“quench” the OD₅₇₀ of hemoglobin was measured by the “Hb” control. These control samples were identical to the hemolysis samples except that a solution of hemoglobin was used in place of intact rbc's. The reduction in OD₅₇₀ caused by each compound in the Hb control is listed in Table 7-4. The Hb control was a reliable way of identifying compounds causing a complete loss of OD due to redox activity (101, 107) but could not always be used to quantitate the percent reduction in OD due to inhibition of fusion.

The results of screening selected derivatives of 83 by hemolysis are shown in Table 7-4 and summarized in Table 7-6. Compounds 83, 91, 97, 117, and 126 apparently inhibited hemolytic activity while 99, 116, 120, and 135 did not. The concentration profiles are in good agreement with those observed by SPA. These data suggest that the certain derivatives of 83 are able to inhibit hemagglutinin-mediated hemolysis.

Figure 7-6 shows hemolysis with compound 91 at 0.1 mM as a function of pH. Maximum inhibition (61.0%) was observed at pH 5.0 and is consistent with the data in Table 7-4. Figure 7-6 also reveals that compound 91 shifts the pH profile of hemolysis by approximately -0.2 pH units.

compound	pure	log conc (M)	hemolysis		Hb control	
			%↓OD	n	%↓OD	n
83		-3	67.0 ± 15.1	4	16.5 ± 2.1	2
		-4	17.9 ± 20.0	2	nd	0
		-5	11.3 ± 10.3	2	nd	0
90	*	-3	90.3	1	68.1	1
		-4	33.1	1	57.1	1
		-5	4.7	1	24.9	1
91		-3	91.0 ± 14.6	4	29.6 ± 2.5	3
		-4	60.5 ± 0.7	2	26.8 ± 5.0	4
		-5	13.2	1	11	1
		-6	-5.1	1	3	1
97		-3	102.0 ± 5.8	4	55.0 ± 13.3	3
		-4	94.6	1	37.6 ± 4.9	2
		-5	23.7	1	31	1
		-6	6.6	1	13.0	1
99	*	-3	28.7	1	27.8	1
101		-3	112.2	1	93.0	1
107		-3	104.1	1	92.5	1
111	*	-3	36.7 ± 0.6	2	64.9 ± 0.5	2
		-4	-6.9 ± 9.3	2	40.4 ± 1.3	2
116	*	-3	17.3	1	44.0	1
		-4	7.7	1	31.6	1
117	*	-3	116.0	1	17.8	1
		-4	31.6 ± 6.9	2	9.9 ± 2.8	2
		-5	13.9	1	7.9	1
120	*	-4	6.2	1	4.4	1
121	*	-3	116.2	1	68.0	1
		-4	31.5 ± 7.8	2	nd	0
		-5	-4.6 ± 0.1	2	nd	0
126	*	-3	101.4 ± 7.6	2	54.4 ± 3.7	2
135	*	-3	80.3	1	75.2	1

Table 7-4. Effect of analogs of compound 83 in hemolysis. The % reduction in OD₅₇₀ of each sample compared to the appropriate DMSO-containing control is given for both hemolysis reactions and Hb controls (see text). Rbcs used in the Hb control for 117 were prelysed in NP40 solution. OD values are averaged over n independent experiments. "nd", not done. '*' indicates a pure or purified compound (Appendix A). Hemolysis Protocol 3 was used.

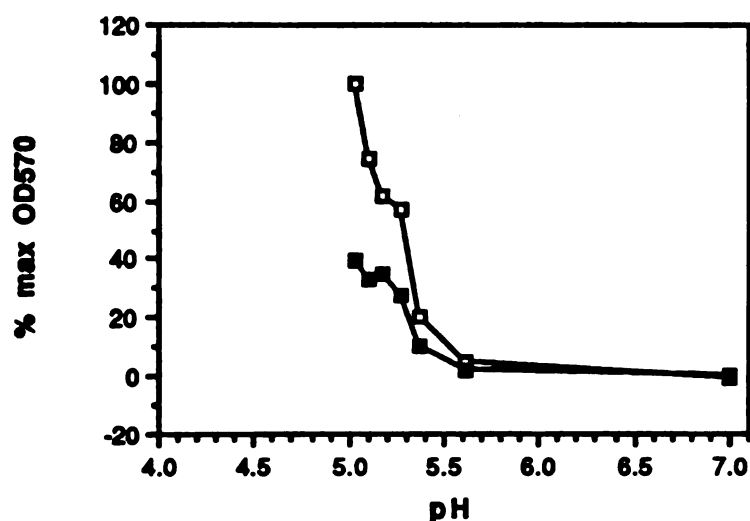


Figure 7-6. Hemolysis with compound 91 as a function of pH. (■) samples containing 0.1 mM 91, (□) samples without 91. Hemolysis Protocol 3 is described in the methods. The OD₅₇₀ of control samples at pH 5.0 and pH 7.0 were taken as 100% and 0%, respectively. Inhibition was not measured below pH 5.0 since hemolysis in control samples at lower pHs is reduced by an unknown mechanism (Figure 5-2).

Fluorescence Dequenching

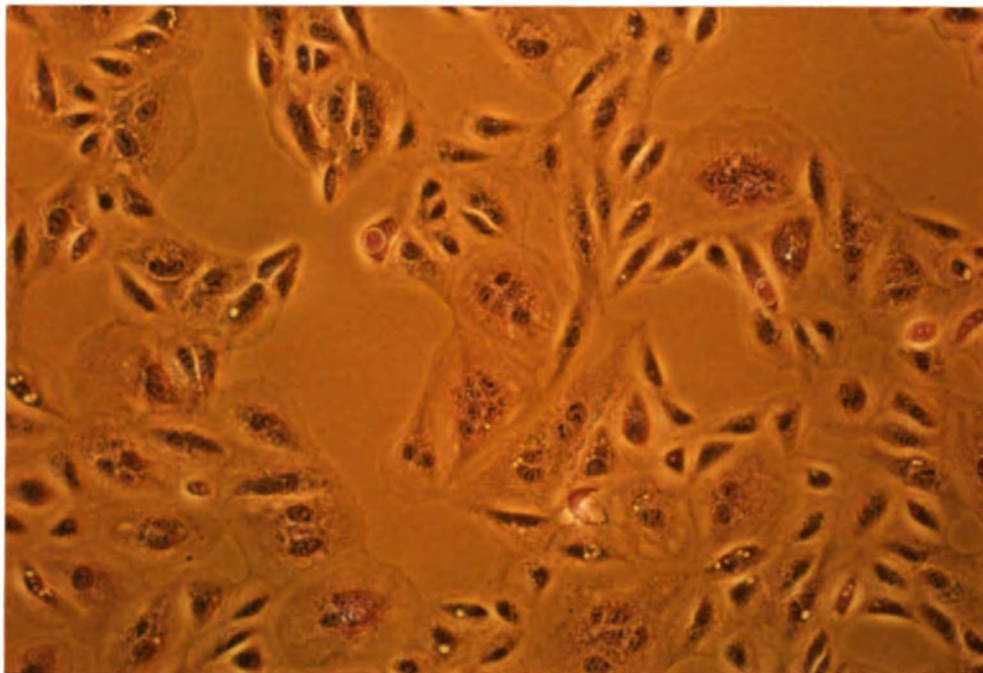
The fluorescence dequenching (FDQ) assay was tested as a means to confirm the ability of the compounds to prevent fusion inhibition. This assay monitors lipid mixing between HA-expressing cells and red blood cells (rbcs) labeled with a fluorescently-tagged lipid. Octadecyl rhodamine (R₁₈) is incorporated into the rbc membranes at self-quenching concentrations. When the probe is diluted into the membrane of an HA-expressing cell, R₁₈ quenching is relieved and an increase of fluorescence is observed.

Unfortunately, this assay could not be used reliably to determine the ability of the compounds to inhibit fusion-dependent lipid mixing. The calculated %FDQ (see Chapter 3) is a function of the baseline fluorescence (F_0), the fluorescence as a function of time at low pH (F_t), and the fluorescence at infinite probe dilution, approximated in a detergent solution (F_{NP40}). Several of the compounds (90, tested at 10 μ M; 91, 0.1 mM; 97, 10 and 100 μ M; 99, 1 mM; 110, 10 and 100 μ M; 117, 1 and 0.1 mM) quenched each of these parameters by unequal amounts. Experiments with rbcS prepared with varying amounts of R₁₈ to approximate the probe dilutions in fusion intermediates showed that quenching is dependent on the density of R₁₈ in the membranes. For example, at 0.1 mM, compound 97 inhibited the quenching of rbcS prepared with one sixth the usual amount of R₁₈ by 72% but that of rbcS prepared with five times more R₁₈ by 86%. Inhibition could not be measured by this assay for compounds 92 (at 1 mM) and 114 (tested at 0.1 mM) which are either fluorescent themselves or relieve self-quenching of R₁₈ by a fusion-independent mechanism. Of the compounds that did not exhibit these problems, none inhibited the FDQ signal at the concentrations tested. These were 91 (10 μ M), 99 (10 or 100 μ M), and 116 (1 or 0.1 mM). These data are in agreement with the SPA and hemolysis data suggesting that, at these concentrations, compounds 91, 99, and 116 do not inhibit fusion.

Fusion from Without

The ability of the compounds to inhibit fusion was also assessed by a third assay. "Fusion from without" measures the ability of viruses bound to cells to mediate syncytia formation. Nontoxic compounds were identified by the MTT assay as described in Chapter 3. However, to minimize misinterpretation of the results, only compounds which did not visibly affect the morphology of the cells were screened by this assay. Figure 7-7 compares syncytia formation in the presence of 117 and 120 to controls with DMSO alone. Significant inhibition was not observed with compound 120 at 10 μ M or 117 at 1 μ M. In contrast, 117 at 10 μ M completely abolished the ability of the virus to induce syncytia. These results are consistent with those from the SPA.

(a)



(b)

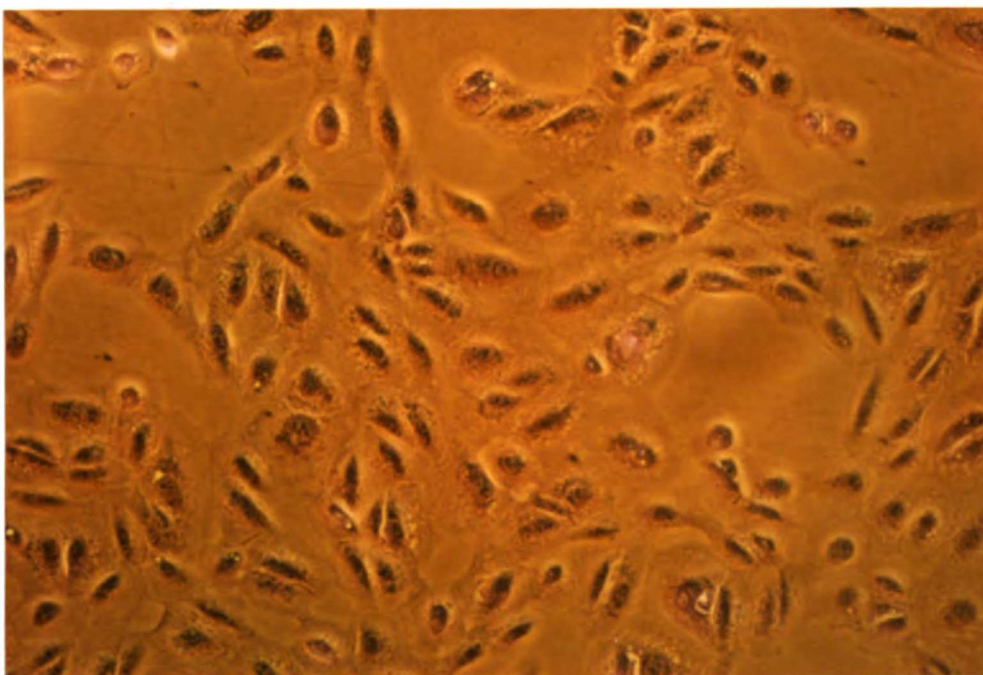
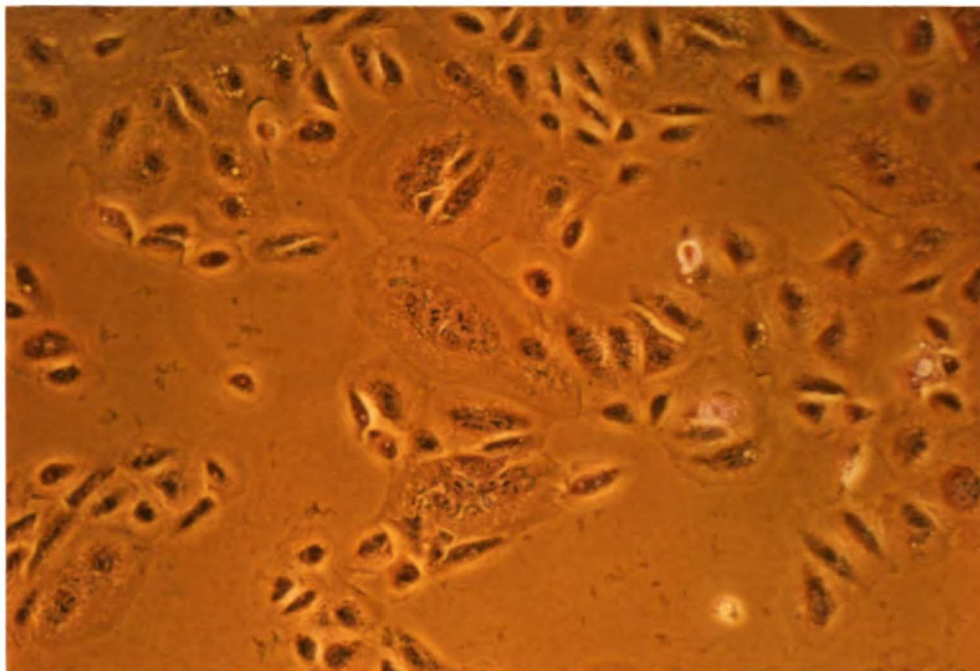


Figure 7-7 continues on the next page.

(c)



(d)

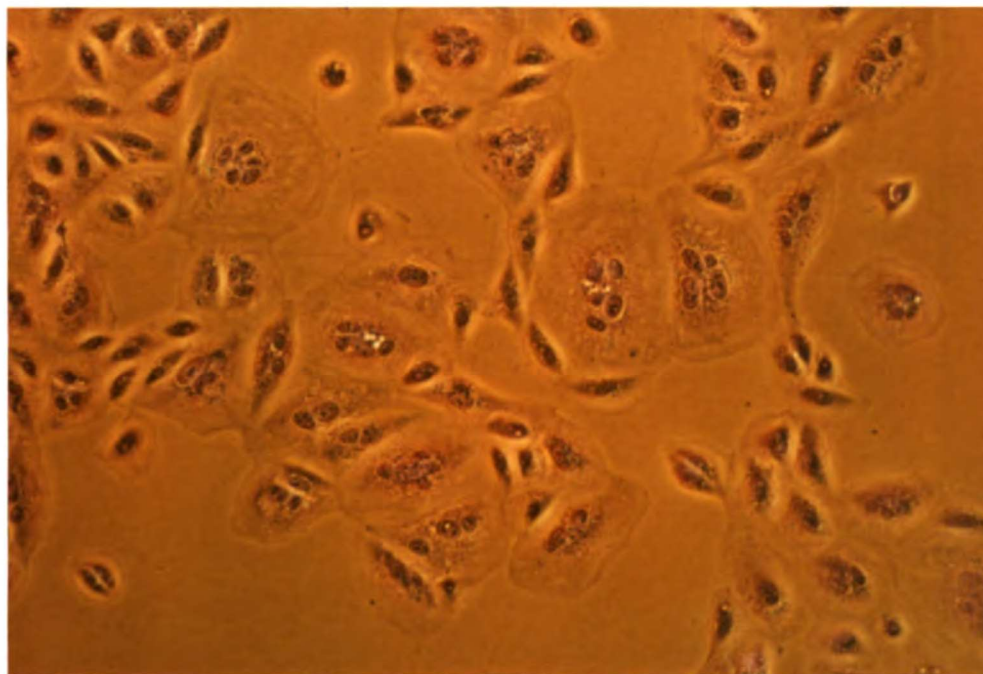


Figure 7-7. Effects of selected compounds on influenza-induced syncytia formation. The “fusion from without” protocol is described in Chapter 3. Wells were incubated with (a) DMSO, (b) 10 μ M 117, (c) 1 μ M 117, (d) 10 μ M 120.

Infectivity

In theory, compounds which inhibit the conformational change and fusion should also prevent viral infectivity. The ELISA-based assay described in Chapters 3 and 5 was used to measure antiviral activity. Cell viability was measured in parallel to distinguish inhibition of viral replication from toxicity. Table 7-5 compares the effects of the compounds in the infectivity and MTT viability assays. While some of the compounds were toxic at the concentrations tested, others were able to inhibit viral replication under conditions at which the cells were viable. IC₅₀s computed by the method described in Chapter 3 are listed in Table 7-6.

compd	pure	log conc (M)	infectivity		viability	
			% inhibition	n	% inhibition	n
83		-3	51.1 ± 28.3	2	92.9 ± 7.4	4
		-4	33.1 ± 20.2	2	46.4 ± 15.1	4
		-5	8.7 ± 6.7	2	7.3 ± 8.9	4
		-6	12.3 ± 2.5	2	3.4 ± 10.5	3
90	*	-4	96.9	1	96.8 ± 3.8	2
		-5	81.7 ± 2.2	2	33.0 ± 9.9	2
		-6	-3.3 ± 2.4	2	-6.7 ± 7.3	2
91		-4	70.9 ± 39.8	2	-4.3 ± 7.0	5
		-5	6.5 ± 7.0	5	-2.0 ± 9.4	5
		-6	-1.3 ± 4.0	2	-0.06 ± 5.1	5
97		-4	nd	0	95.9 ± 8.0	4
		-5	92.6 ± 12.7	2	32.9 ± 30.9	9
		-6	-4.7 ± 9.4	2	-3.1 ± 7.9	9
99	*	-4	113.7 ± 12.0	2	99.4 ± 0.4	2
		-5	35.9 ± 5.2	2	-10.1 ± 10.0	2
		-6	26.2 ± 8.6	2	-4.2 ± 9.0	2
101		-4	nd	0	97.4	1
105		-4	nd	0	100.5	1
107		-4	nd	0	96.9	1
110		-3	101.9	1	96.9 ± 0.4	2
		-4	91.9 ± 11.5	2	98.3 ± 1.2	6
		-5	61.1 ± 50.8	8	19.2 ± 15.6	8
		-6	-9.8 ± 0.6	2	5.0 ± 12.5	7
111	*	-4	102.4 ± 10.6	2	41.0 ± 17.0	2
		-5	57.4 ± 25.2	2	-1.6 ± 0.8	2
		-6	21.7 ± 4.9	2	1.0 ± 5.8	2
114	*	-4	nd	0	102.2	1
		-5	92.3 ± 12.7	2	53.7 ± 2.1	2
		-6	0.2 ± 0.2	2	-4.3 ± 3.3	2
116	*	-4	78.1 ± 10.0	2	-23.6 ± 5.0	2
		-5	42.8 ± 4.7	2	-9.4 ± 5.0	2
		-6	18.8 ± 1.7	2	3.6 ± 4.5	2
117	*	-4	53.7 ± 26.4	2	-26.1 ± 3.3	2
		-5	22.2 ± 13.2	2	-13.7 ± 2.8	2
		-6	7.5 ± 0.1	2	-8.4 ± 7.6	2
119		-4	nd	0	97.8	1
120	*	-4	-2.3 ± 3.1	2	-7.0 ± 4.4	2
		-5	4.7 ± 0.9	2	-6.0 ± 5.0	2
		-6	7.1 ± 6.7	2	0.5 ± 0.3	2
121		-4	nd	0	100.3	1
124		-4	nd	0	80.5	1
126	*	-4	35.6 ± 0.6	2	-7.6 ± 7.3	2
		-5	7.4 ± 1.8	2	-9.2 ± 0.6	2
		-6	3.8 ± 5.5	2	-6.9 ± 4.0	2

Table 7-5. Effect of the analogs of compound 83 on infectivity and cell viability. Experimental protocols and calculation of % inhibition

are described in Chapter 3. All samples were done in triplicate in each experiment. The number of experiments averaged, n, is listed with the mean inhibition and standard deviation. Untested compounds ("nd") are completely toxic at concentrations at which specific inhibition is expected. '*' indicates a purified ligand (Appendix A).

Summary of Experimental Results

compound	pure	IC ₅₀			
		SPA	hemolysis	syncytia	infectivity
83K	*	10 ⁻³ -10 ⁻⁴	10 ⁻³ -10 ⁻⁴	nd	toxic
83A	*	10 ⁻⁴ -10 ⁻⁵	--†	nd	--†
84		10 ⁻³	nd	nd	nd
90	*	10 ⁻³ -10 ⁻⁴	10 ⁻³	morph	10 ⁻⁵
99	*	>10 ⁻³	>10 ⁻³	toxic	>10 ⁻⁵
105		≥10 ⁻³ -10 ⁻⁴	nd	nd	toxic
107		≥10 ⁻³ -10 ⁻⁴	quench	nd	toxic
111	*	10 ⁻⁵	10 ⁻³ -10 ⁻⁴	morph	10 ⁻⁵
112		≥10 ⁻³ -10 ⁻⁴	nd	nd	nd
114	*	>10 ⁻⁵	nd	toxic	>10 ⁻⁵
115		≥10 ⁻³ -10 ⁻⁴	nd	nd	nd
116	*	>10 ⁻³	>10 ⁻³	morph	10 ⁻⁴ -10 ⁻⁵
117	*	10 ⁻⁵ -10 ⁻⁶	10 ⁻⁴	10 ⁻⁵ -10 ⁻⁶	10 ⁻⁴ -10 ⁻⁵
120	*	>10 ⁻³	>10 ⁻⁴	>10 ⁻⁵	>10 ⁻³
121	*	>10 ⁻³	≥10 ⁻³	nd	toxic
124		>10 ⁻³	nd	nd	toxic
126	*	10 ⁻³	10 ⁻³	nd	>10 ⁻⁴
135	*	>10 ⁻³	>10 ⁻³	nd	nd

Table 7-6. Summary and comparison of results from the scintillation proximity, hemolysis, syncytia, and infectivity assays. The midpoint of the dose response curves for each compound in each experiment is

shown. Inhibitory activity has been corrected for nonspecific effects (quenching in SPA, hemoglobin oxidation in hemolysis, and cell toxicity in infectivity) as described in Chapter 3. A range of concentrations indicates that the IC_{50} fell between two tested concentrations. For example, 10^{-3} - 10^{-4} implies that over 50% inhibition was observed at 1 mM and but less than 50% inhibition was observed at 0.1 mM. $>10^{-3}$ indicates that 50% inhibition was not achieved at 1 mM, the highest concentration tested. Data for compounds without POST controls in the SPA (represented by \geq) are included as an upper bound on the activity of the sample. "quench" means that inhibitory activity could not be assessed due to quenching by the compound. "toxic" represents compounds determined to be toxic to the cells used in the assay. "morph", the compound altered the morphology of the cells. "nd", not done. "*" indicates a purified compound (Appendix A). 91, 97, 125 are impure, 110 is unstable, and 93, 113, 123 are insoluble in DMSO. Compounds 85, 86, 87, 88, 89, 92, 94, 95, 96, 98, 100, 101, 102, 103, 104, 106, 108, 109, 118, 119, 122, and 127 showed no inhibition in the SPA assay at 1 μ M - 1 mM. Note that hemagglutinin is exposed to low pH for 5 minutes in the SPA and syncytia assays but for 15 minutes in hemolysis.

† Compound 83A was initially active in the SPA assay but lost activity as it aged. No active sample was available for testing in the other assays.

Table 7-6 compares the results of screening compound 83 and its derivatives in conformational change, hemolysis, syncytia, and infectivity assays. Due to nonspecific effects, measurement of inhibition of fusion or infectivity was impossible for a number of the compounds. The available data show that compounds which inhibited the SPA also inhibited hemolysis, fusion from without, and infectivity. Although 116 was inactive in the SPA but inhibited infectivity, no other active compound showed inhibitory activity in any assay.

Among the active compounds, there is generally good agreement between the concentration profiles in each assay. Results from compounds 90, 11, and 117 show that hemolysis consistently required somewhat higher concentrations to achieve 50% inhibition. Although hemolysis, syncytia formation, and infectivity can be inhibited in many ways, the coincidence of the observed concentration profiles with those from SPA analysis suggests that at least part of the observed inhibition has a common mechanism. Mounting evidence supports the idea that the interactions causing inhibition are specific for hemagglutinin and not proteins in general. The compounds do not abolish the binding capacity of antibodies, protein A, or the viral receptor on rbc's, they are not lethal to cells at all active concentrations, and they do not inhibit proteinase K (not shown) or HIV-1 protease [2]. Combined with the observed reversibility, the data imply that the compounds are exerting a noncovalent effect on hemagglutinin resulting in inhibition of the fusion-inducing conformational change.

The two most active compounds, 111 and 117, are structurally related. Comparison of the structures of these compounds with less active or inactive compounds suggests which structural features improve activity and which reduce inhibition. This model is shown in Figure 8-1 and discussed in Chapter 8.

Table 7-7. Commercial and synthesized derivatives of 83.

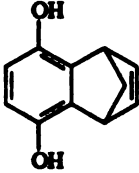
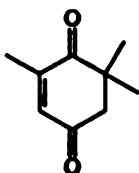
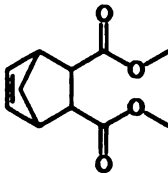
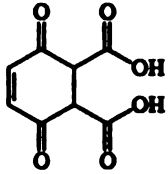
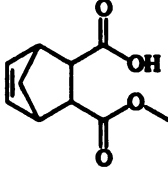
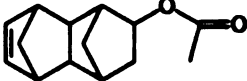
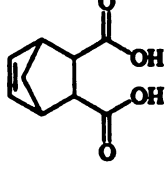
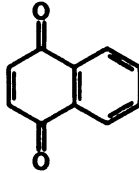
ID	name	source	FW	structure
83A	5,8-dihydro-5,8-methano-1,4-naphthalenediol	SERES	174	
84	2,6,6-trimethyl-2-cyclohexene-1,4-dione	Fluka	152	
85	dimethyl-bicyclo[2.2.1]-5-heptene-2,3-dicarboxylate	Lancaster Synthesis	210	
86	3,6-endoxo-1,2,3,6-tetrahydrophthalic acid	Tokyo Kasei	184	
87	5-norbornene-2,3-dicarboxylic acid monomethyl ester	TCI	196	
88	1,2,3,4,4A,5,8,8A-octahydro-1,4:5,8-dimethanonaphth-2-yl acetate	Bader	218	
89	cis-5-norbornene-endo-2,3-dicarboxylic acid	Aldrich	182	
90	1,4-naphthoquinone	Aldrich	158	

Table 7-7, continued.

91	5,8-dioxo-1,4,4A,5,8,8A-hexahydro-1-naphthalene carboxylic acid	Bader	206	
92	5,5-dichlorotetracyclo(6.2.1.0.2,7.0.4.6)undec-9-ene	Bader	215	
93	1,4,4A,4B,5,8,8A,8B,9,10-decahydro-1,4:5,8-dimethanoanthracene-9,10-dione	Bader	240	
94	N-hydroxy-5-norbornene-2,3-dicarboximide	Aldrich	179	
95	N-methyl-5-norbornene-2,3-dicarboximide	Bader	177	
96	1,2,3,4,4A,5,8,8A-octahydro-1,4:5,8-dimethano-2-naphthol	Bader	176	
97	methyl 5,8-dioxo-1,4,4A,5,8,8A-hexahydro-1-naphthalene carboxylate	Bader	220	
98	3-(N-phenylcarbamoyl)-5-norbornen-2-carboxylic acid	Bader	257	

Table 7-7, continued.

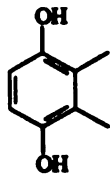
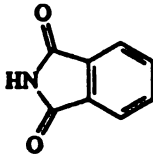
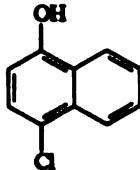
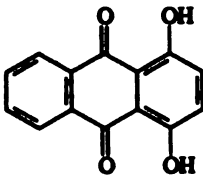
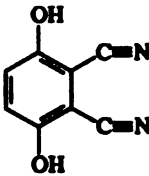

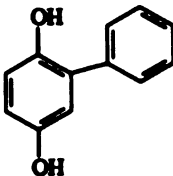
99	2,3-dimethyl-hydroquinone	Aldrich	138	
100	phthalimide	Aldrich	147	
101	4-chloro-1-naphthol	Aldrich	179	
102	quinizarin	Aldrich	240	
103	2,3-dicyanohydroquinone	Aldrich	160	
104	hydroquinone	Aldrich	110	
105	phenyl hydroquinone	Aldrich	186	

Table 7-7, continued.

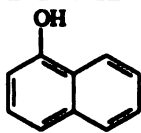
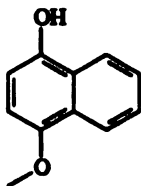
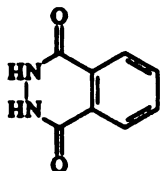
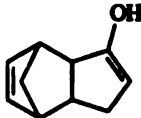
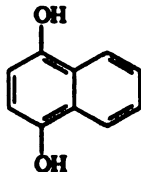
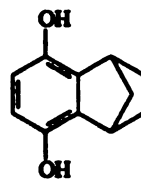
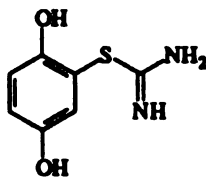
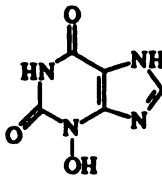
106	1-naphthol	Aldrich	144	
107	4-methoxy-1-naphthol	Aldrich	174	
108	phthalhydrazide	Aldrich	162	
109	hydroxydicyclopentadiene	TCI	150	
110	1,4-dihydroxynaphthalene	TCI	160	
111	3',6'-dihydroxybenzonorbornane	TCI	176	
112	2-(2,5-dihydroxyphenyl)-2-thiopseudourea hydrochloride	Bader	221	
113	xanthine 3-N-oxide	Sigma	168	

Table 7-7, continued.

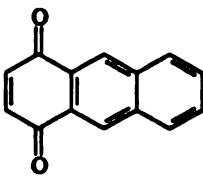
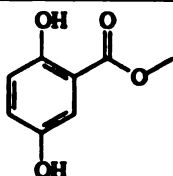
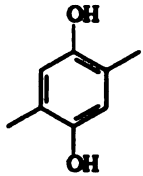
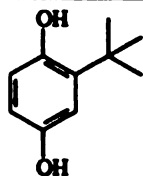
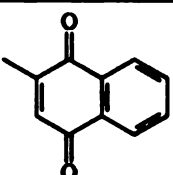
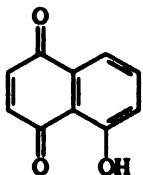
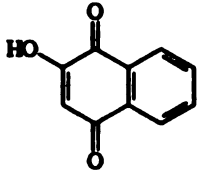
114	1,4-anthraquinone	Lancaster Synthesis	208	
115	methyl-2,5-dihydroxybenzoate	Frinton Labs	168	
116	2,5-dimethylhydroquinone	Pfaltz & Bauer	138	
117	tert-butyl-hydroquinone	Aldrich	166	
118	2-methyl-1,4-naphthoquinone	Aldrich	172	
119	5-hydroxy-1,4-naphthoquinone	Aldrich	174	
120	2-hydroxy-1,4-naphthoquinone	Aldrich	174	

Table 7-7, continued.

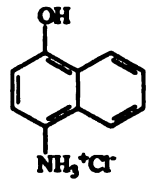
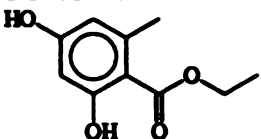
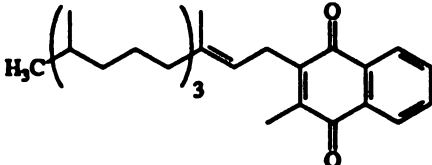
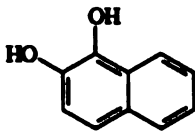
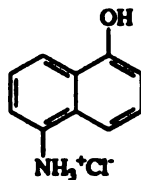
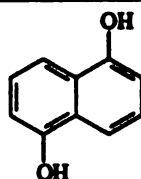
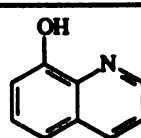
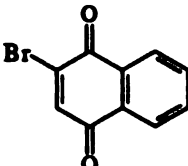
121	4-amino-1-naphthol hydrochloride	Aldrich	196	
122	ethyl-2,4-dihydroxy-6-methylbenzoate	Aldrich	196	
123	vitamin K1	Aldrich	451	
124	1,2-dihydroxynaphthalene	Aldrich	160	
125	1-amino-5-naphthol	Pfaltz & Bauer	161	
126	1,5-naphthalenediol	Aldrich	162	
127	8-hydroxyquinoline	Baker	147	
135	2-bromo-1,4-naphthoquinone	SERES	237	

Table 7-7 notes. ID number, name, and formula weight are listed for each compound. The company which either sold or synthesized each compound is given in addition to the chemical structure.

Crystallography

Although the results of the biological assays suggest that the compounds interact with hemagglutinin, they do not prove that the compounds are actually binding to that protein or, if they are, whether they are binding to the targeted site. To address these questions, we have begun a collaborative project with Dr. Don Wiley's laboratory at Harvard to determine the crystal structures of complexes of the active compounds and BHA.

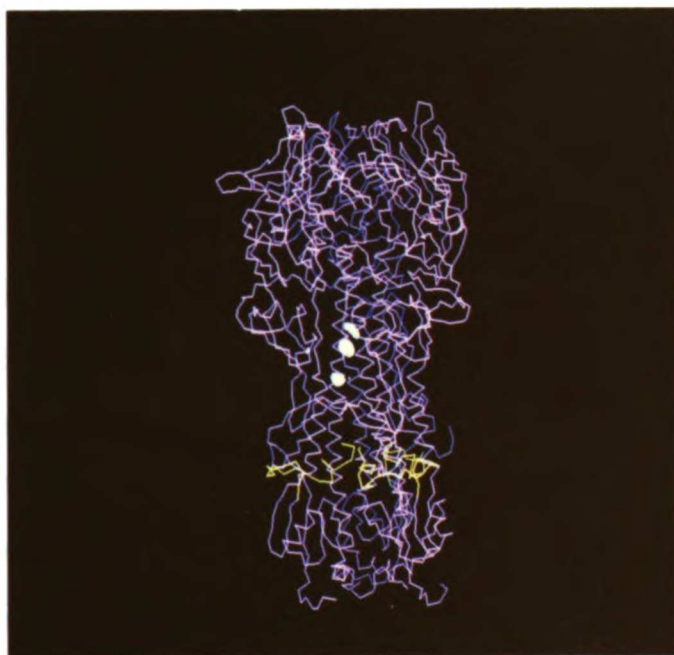


Figure 7-8. Location of the difference density (white) from the 90-BHA co-crystal superimposed on the trimer. The 2.5σ contours

including the major (5.6σ) peak are shown for the hinge region of the A/B monomer. The α -carbon backbone (magenta) of BHA is displayed with the fusion peptides colored yellow.

Compounds 90, 99, and DMSO have been soaked into BHA crystals, and three-fold averaged difference maps from 5.2 \AA data have been computed. A 5.6σ peak was observed in the map from the 90-containing crystal, with a much smaller one in the same position from the co-crystal with 99. There is no peak present in the difference map of BHA with DMSO alone. The density lies in the hinge region near Trp 92 of HA2, at the interface between monomers (Figure 7-8). Attempts are underway to reproduce these results and to obtain co-crystals with compound 117. Strategies are also being considered to test whether the location of the difference density represents the source of the observed biological activity.

DOCKing to the Crystallographic Site

DOCK3.0 was used to dock the uncharged³ derivatives of 83 into the site identified by the crystallography. This version of DOCK scores orientations with an AMBER-derived force field which evaluates both electrostatic and van der Waals interactions [3]. Parameters used in the docking are listed in Table 7-8 and the resulting scores for the active compounds in Table 7-9. Figure 7-9a plots the force field score of each compound against the approximate

³Only uncharged derivatives were considered since the potential function tends to overestimate electrostatic contributions.

IC₅₀ derived by SPA. DOCK is unable to discriminate the active from inactive compounds, nor predict the relative affinities of the inhibitors. For discussion, see Chapter 8.

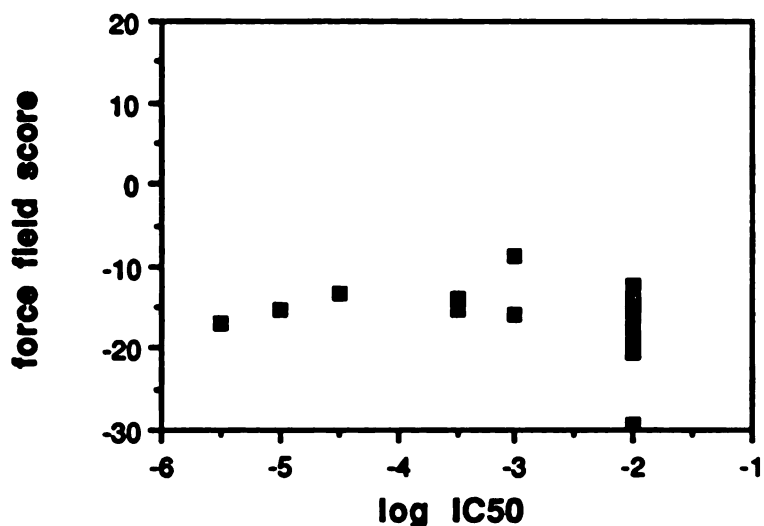
variable	value
<i>dislim</i>	1.5
<i>nodlim</i>	5
<i>ratiom</i>	0
<i>lownod</i>	4
<i>lbinsz</i>	0.1
<i>lovlap</i>	0.0
<i>sbinsz</i>	1.0
<i>sovlap</i>	1.0
<i>nbump</i>	0
<i>expmax</i>	0
<i>intrp</i>	0
<i>vdwmax</i>	1000000.
<i>escale</i>	1.
<i>vscale</i>	1.

Table 7-8. Parameter values input to DOCK3.0.

compound	log IC ₅₀	force field score	
		crystallographic site	tail site
83K	-3.5	-13.99	-13.65
83A	-4.5	-13.41	-12.86
84	-3	-8.77	5.75
90	-3.5	-15.34	-7.35
111	-5	-15.44	-9.46
117	-5.5	-16.88	-15.68
126	-3	-15.81	-12.11

Table 7-9. Scores of active compounds. Ligands were docked to the site identified by the crystallography and to the tailA site with DOCK3.0. Parameter values are listed in Table 7-8. More negative scores suggest increased complementarity to the site. IC₅₀'s are estimated from SPA data.

(a)



(b)

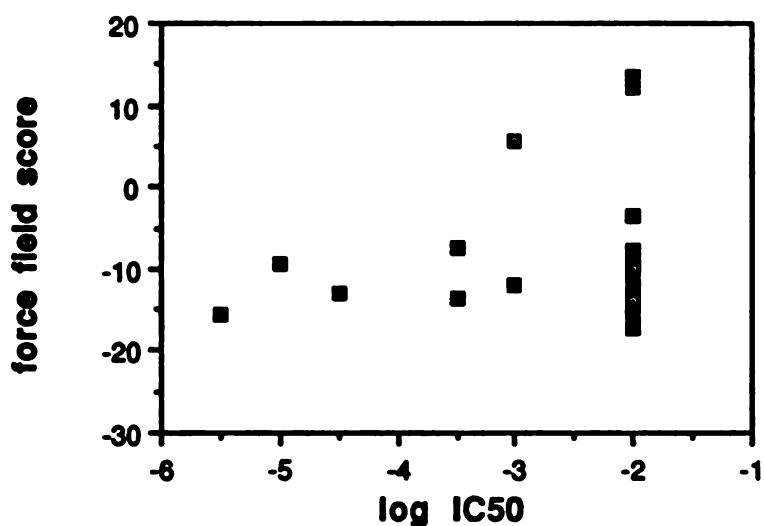


Figure 7-9. Comparison of the IC₅₀ and force field score of each derivative of 83 when docked to (a) the site with crystallographic difference density or (b) the tail1A site. Only uncharged ligands were considered. Compounds with more extensive complementary to the receptor site are represented by more negative scores. IC₅₀s are listed in Table 7-6. Inactive compounds are assigned an IC₅₀ of -2. Compounds with IC₅₀s falling between two tested concentrations are graphed midway between the bracketing values. Input parameters to DOCK3.0 are listed in Table 7-8.

Figure 7-9b compares the results obtained from docking the same compounds to the original tail site. In this case, DOCK slightly discriminates the more active compounds from those less active. The highest scores, indicative of poor complementarity, were assigned to inactive compounds. Interestingly, compound 117, the most active analog to date, ranked fifth when docked to this site and was assigned the best score of all inhibitory compounds when docked to this site or the crystallographic site.

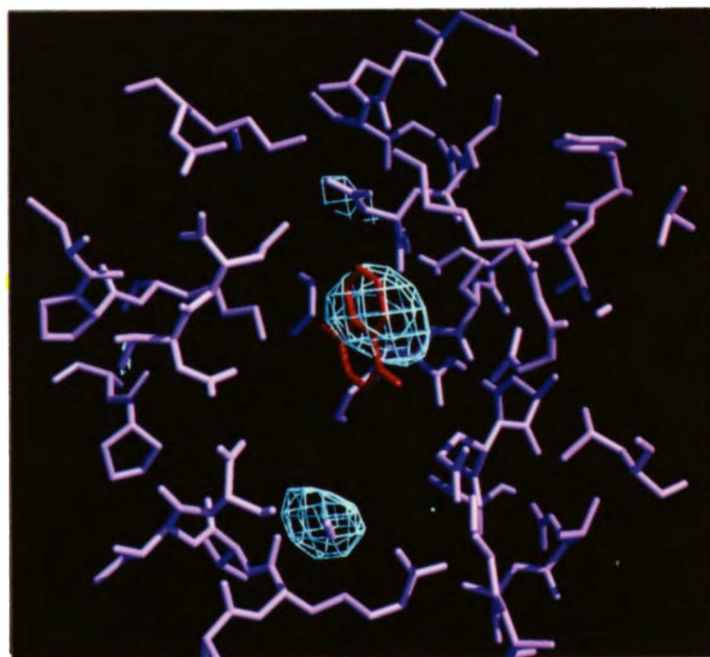


Figure 7-10. Superposition of the docked configuration of compound 90 on the difference density from the 90-BHA complex. The 2.5 σ contours (blue) are shown in the vicinity of the site (magenta). The major peak coincides with the dock orientation receiving the best force field score (red).

While DOCK did not correctly rank the compounds by their activity it did superimpose compound 90 over the major peak of the difference density (Figure 7-10). This site has a shallow rim with a deep, tight pocket on the bottom. The shape scoring schemes of DOCK1 and DOCK2 orient all the compounds into this snug cavity. In contrast, DOCK3 placed most of the derivatives over one of the two largest peaks of difference density. Comparison of the modeled configurations to the precise orientation of the compounds within the subsite awaits determination of a high resolution structure.

Summary

The Round II compounds were screened for inhibition of fusion peptide exposure using the SPA assay. One apparently reversible inhibitor, compound 83, was found. Inhibition was observed with analogs of 83 in the SPA, hemolysis, syncytia, and infectivity assays. The concentration profile of each compound was consistent between the various experiments, suggesting a common mechanism for inhibition of the conformational change, fusion, and infectivity. Preliminary crystallographic results with one of the derivatives suggest this compound is binding to hemagglutinin. However, the putative binding site is not the "tail1A" site targeted in the modeling but a site in the hinge region of the trimer. The analogs of 83 were DOCKed to both the tail1A site and the crystallographic site but no discrimination was observed.

References

1. Martin, K. and Helenius, A., "Transport of incoming influenza virus nucleocapsids into the nucleus." *J Virol* 65 232-244 (1991).
2. Babe', L., personal communication.
3. Meng, E. C., Shoichet, B. K. and Kuntz, I. D., "Automated docking with grid-based energy evaluation." *J Comp Chem* 13 505-524 (1992).

Chapter 8

Discussion

Infection of a cell by enveloped viruses requires fusion between the membrane of the virus and a target cell membrane. In the case of influenza, fusion is mediated by hemagglutinin, a trimeric glycoprotein integral to the viral membrane [1, 2]. The fusion peptide, a hydrophobic segment directly involved in the fusion reaction, is buried in the trimer interface of native hemagglutinin [3, 4]. Fusion is triggered by a pH-dependent conformational change in hemagglutinin that releases the fusion peptide from its unexposed position [5, 6]. Since the fusion-inducing conformational change is crucial for viral replication, inhibition of fusion peptide release should prevent viral infection.

Identification of inhibitors

Structural and biological information about hemagglutinin has been used in conjunction with the inhibitor design method, DOCK, to identify inhibitors of the conformational change. Antiviral compounds with this activity have been found in a family of small quinones and hydroquinones (structures depicted in Table 7-7). The strongest derivative to date has an IC_{50} between 1 and 10 μM (Table 7-6). Representative compounds also inhibit hemagglutinin-

mediated hemolysis and influenza-induced syncytia formation (Table 7-4 and Figure 7-7).

Figure 8-1 summarizes the structural features shared among the active compounds. The strongest inhibitors are 1,4 hydroquinone derivatives with a single hydrophobic ring fused at the 5-6 bond or a hydrophobic substituent smaller than a phenyl ring at the 6 position. The corresponding quinones showed reduced inhibition. Derivatization at both C5 and C6 (other than a fused ring) or at C2 also reduced activity.

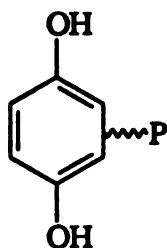


Figure 8-1. Model of the structural features shared among the active compounds, based on the SPA results of compounds 83-135. "P" represents a single hydrophobic ring fused at the 5-6 bond or a hydrophobic substituent smaller than a phenyl ring at the 6 position.

This information will guide the design of compounds predicted to be more effective inhibitors. As data from screening additional derivatives become available the model in Figure 8-1 will be refined and tested with ligands designed to reflect the changes. High resolution structural information will also be used to predict features likely to improve binding. Detailed analysis of the interactions

between the compounds and the protein will suggest other compounds, perhaps based on alternate chemical frameworks, that are likely to have increased steric and electrostatic complementarity to the binding site(s). Any compound with sufficiently high activity and specificity will be tested for its pharmacologic properties and perhaps lead to the development of an anti-influenza drug.

The compounds will also be screened for antiviral activity against additional strains of influenza as well as other enveloped viruses. Syncytial assays already underway reveal that some of the compounds are able to inhibit fusion mediated by the HIV-1 fusion protein but with different IC_{50} s from those observed with influenza [7]. Compound 90 was active at 2 μ M and 107 at 50 μ M. No inhibition beyond toxicity was observed with compounds 83K at >100 μ M or 99 at 10 μ M.

Besides being novel leads for drug design, these compounds can also serve as probes of the conformational change of hemagglutinin. The site identified from the co-crystal with compound 90 lies ~30 Å from the fusion peptide. If the crystallographic binding site does regulate fusion peptide exposure, it suggests that the conformational change is cooperative throughout much of the trimer. A similar finding has been reported recently from studies of a mutant hemagglutinin with disulfide bonds engineered ~75 Å from the fusion peptide region [8, 9]. This mutant is unable to change conformation or mediate fusion. Protease digestion studies and immunoprecipitations with other antibodies will reveal which other parts of the conformational change are inhibited in the mutant and by the compounds.

Evaluation of the biological strategy

To date, inhibition of fusion has not been widely tested as a route for antiviral development [10]. Since viral fusion is not an enzymatic process, inhibitor discovery methods requiring knowledge of substrates, inhibitors, and/or catalytic activity can not be applied. Previous fusion inhibitors that interact with viral proteins have been discovered serendipitously [11, 12]. Therefore, a structure-based design method which only requires knowledge of the receptor site was used to identify antifusogenic compounds targeted against hemagglutinin. The apparent ability of compound 83 and its derivatives to reduce fusion peptide exposure, fusion, and infectivity of influenza demonstrates that inhibition of the conformational change of the fusion protein is a realizable strategy for inhibiting viral infection. These compounds represent the first antiviral agents intentionally designed to inhibit the vital conformational changes of any enveloped (or nonenveloped) virus. As three-dimensional coordinates of other fusion proteins become available from biophysical experiments or from modeling, *de novo* fusion inhibitors can be designed against an increasing number of viruses.

Studies of the detailed mechanisms of the conformational changes of fusion proteins will help identify binding sites likely to regulate fusion activity. In the case of influenza hemagglutinin, recent studies suggest that antiviral agents can be targeted not only to the vicinity of the fusion peptide itself but to sites quite distant from it (see Chapter 9) [8, 9]. Since the conformational change is irreversible, an alternate route to fusion inhibition is to prematurely

trigger fusion peptide exposure and thus inactivate hemagglutinin [13, 14]. Recent studies suggest that this is part of the mode of action of the drug amantadine [15, 16].

Conformational changes are ubiquitous among both normal and disease processes. Thus, inhibition of conformational changes is a strategy applicable to the control of a wide range of functions, from intoxication by bacterial toxins [17] to metabolism [18]. The inhibitors can take a variety of forms. The WIN compounds and arildone are drugs that stabilize the viral capsids of rhinoviruses and polioviruses and prevent uncoating [19, 20, 21]. Antibodies can also inhibit the conformational changes required for uncoating [22]. Covalent inhibitors that prevent the transition between the oxy- and deoxy- forms of hemoglobin have been identified by structure-based design [23, 24].

Evaluation of DOCK

The structure-based design method DOCK was used to identify a site on hemagglutinin that could theoretically regulate fusion peptide exposure and to select potential inhibitors of the conformational change. Although the inhibitory compounds may be binding elsewhere, the "tail1A" site over histidine 17 of HA1 is an attractive target for anti-influenza drug design efforts. The high conservation of the site among various strains of influenza (Table 4-5) and its proximity to the fusion peptide suggest an important role for the proper functioning of hemagglutinin. Several mutant hemagglutinins which fuse at elevated pH have substitutions at

position 17 [25, 26]. While it is unknown whether the mutations alter hydrogen bonding patterns, site geometry, or the electrostatic environment, they suggest that residues comprising the site do have a role in regulating fusion peptide exposure.

Use of DOCK led to the discovery of an inhibitor with the desired activity. An active agent was identified after testing only 78 compounds, compared to the thousands typically tested in random screening trials [27]. Although the “rationality” of the discovery is questioned by the preliminary crystallographic data, it is intriguing that in every test case to date, the DOCK method has identified inhibitors despite errors in the predicted binding modes [28, 29, 30].

Hemagglutinin was a particularly difficult test for DOCK. Previous work has shown that reliable dockings require crystal structures with resolution 2.5 Å or better [31], but the hemagglutinin structures available were solved to only 3 Å resolution [4, 32]. Successful predictions for this test case demand not only that the compounds bind to the targeted site, but that the newly discovered pocket possess the hypothesized activity. If the site does not regulate fusion peptide exposure, compounds which do bind there will be overlooked since the screening assays measure inhibition.

The hemagglutinin project progressed with the evolution of DOCK and thus tests the effectiveness of the changes in the method in leading to inhibitor discovery. The inactive Round I compounds were selected with DOCK1 and the Round II compounds including the active 83 were identified with DOCK2. Modifications between DOCK versions 1 and 2 include changes in the databases searched, the

generation of ligand configurations, and the scoring function [33]. Each of these changes improved the program to varying degrees.

The databases searched with DOCK1, derived from the Cambridge Crystallographic Database (CCD), contain experimentally determined coordinates. In contrast, the structures of the Fine Chemicals Directory (FCD) were calculated from the connectivity of each ligand. Initial reservations about using computer-generated coordinates proved unfounded since the calculated structures are generally reasonable [34]. The larger problem stems from docking ligands as rigid bodies, whether using experimental or computed coordinates. In order for DOCK to reproduce configurations of molecules known to bind to a target site the conformation of the ligand in the database must approximate the bound conformation. This problem has contributed to the inaccuracy of at least one prediction [29]. Methods that allow ligands conformational flexibility are being developed [35, 36, 37]. The converse issue of conformational changes in the receptor is a more difficult problem.

When CCD-derived databases were used as a source of potential inhibitors, the designed compounds generally required synthesis. The number of structures that could be tested experimentally was limited by the throughput of the synthetic chemist and the variety was limited by the need for synthetic accessibility and the biases of the chemist. In contrast, searching the FCD of commercially available compounds removed the dependence on organic synthesis and thus permitted selection of a greater number and variety of chemicals. In either search, the diversity of compounds that could be selected was

limited by the contents of the database but can be increased by expanding the database or by allowing design of new structures [38].

Both the CCD and FCD contain tens of thousands of chemicals. To reduce the time required to dock every compound at high resolution, an initial coarse search was done to identify classes of compounds most likely to bind to the target site. Members of each selected class were then examined in more detail. As the databases grow larger, the idea of clustering the compounds becomes more attractive. If the database is pre-sorted into sets of structurally and/or chemically related molecules, then the decision to dock the individual members of the group can be based on the results from representative compound(s). Entire sets of chemicals predicted not to bind can be eliminated by docking one (or a few) compounds. Methods for clustering databases are being developed [39, 40].

Other changes between DOCK1 and DOCK2 were algorithmic. Modifications in the generation of ligand orientations and in the site definition allow higher resolution searches [33]. Because more configurations are tested, DOCK2 is more likely to find the best orientation for each ligand than DOCK1, and thus is less likely to overlook a good candidate.

However, even if the correct orientation of a ligand is generated, it has to be recognized as such. The scoring function must identify the single orientation most likely to approximate the geometry of the ligand when bound to the site. DOCK1 scores attractive interactions between ligand and receptor atoms as a continuous function of the distance between them [41]. In DOCK2, the score is incremented by an integral amount for each distance

within the specified range [33]. Although discretization inherently reduces resolution, the scoring functions are themselves of limited accuracy and the faster scheme of DOCK2 enables examination of more orientations and more compounds.

The scoring schemes in DOCK1 and DOCK2 only evaluate steric complementarity. This shortcoming is partially redressed in DOCK2 by differentiating receptor atoms by their polarity [33]. Electrostatic interactions are evaluated independently in DOCK2.1 [31]. In DOCK3, the scoring function is derived from the AMBER force field equation and includes both electrostatic and van der Waals terms [42, 43]. DOCK3 is able to reproduce the crystallographic configurations of a variety of ligand-receptor complexes. Scoring schemes that evaluate other factors that contribute to binding, e.g. entropy, are under development [31].

In addition to predicting binding geometry, the scoring scheme should also rank compounds by their affinity for the site. In the hemagglutinin test case, DOCK was not able to differentiate the active from inactive compounds, nor correctly rank the active ligands (Chapter 7). However, there are three caveats to this conclusion. The receptor site used in the docking may not be the actual binding site of the inhibitors. The data are of limited accuracy due to the intrinsic variability of the biological experiments. And the DOCK results were compared to the inhibitory activity of the compounds rather than their binding affinity.

Nonetheless, the hemagglutinin test case points to the need to consider alternate binding sites. Accurate predictions, particularly for large receptors with many potential binding sites, require

evaluation of differential affinity. However, binding site preferences determined by small differences in energy may not be accurately assessed with the current DOCK scoring schemes.

In conclusion, this project both tested a strategy and introduced a family of compounds that could lead to the development of anti-influenza and other antiviral drugs. A previously unrecognized site on hemagglutinin was found that could prove useful for future inhibitor design efforts and in studies of the mechanism of the conformational change. Although the DOCK method identified inhibitors at a higher rate than expected by random screening, there are still many improvements needed. The main values of computer-assisted design in this instance were to aid the selection of a variety of structures and to eliminate compounds unlikely to bind. Before *de novo* design methods can be consistently reliable, the fundamental basis of what makes one molecule bind to another must be better understood and the forces known to be important must be more accurately represented.

References

1. White, J., Helenius, A. and Gething, M.-J., "Hemagglutinin of influenza virus expressed from a cloned gene promotes membrane fusion." *Nature* 300 658-659 (1982).

2. Wiley, D. C. and Skehel, J. J., "The structure and function of the hemagglutinin membrane glycoprotein of the influenza virus." *Ann Rev Biochem* 56 365-394 (1987).
3. Stegmann, T., Delfino, J. M., Richards, F. M. and Helenius, A., "The HA2 subunit of influenza hemagglutinin inserts into the target membrane prior to fusion." *JBC* 266 18404-18410 (1991).
4. Wilson, I. A., Skehel, J. J. and Wiley, D. C., "Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution." *Nature* 289 366-373 (1981).
5. White, J. M. and Wilson, I. A., "Anti-peptide antibodies detect steps in a protein conformational change: low pH activation of the influenza virus hemagglutinin." *J Cell Bio* 105 2887-2896 (1987).
6. Doms, R. W., White, J., Boulay, F. and Helenius, A., "Influenza virus hemagglutinin and membrane fusion." In J. Wilschut and D. Hoekstra (Eds.), Membrane Fusion. (Dekker, New York, 1991) pp. 313-336.
7. Babe', L., personal communication.
8. Godley, L., Pfeifer, J., Steinhauer, D., Ely, B., Shaw, G., Kaufmann, R., Suchanek, E., Pabo, C., et al., "Introduction of intersubunit disulfide bonds in the membrane-distal region of the influenza hemagglutinin abolishes membrane fusion activity." *Cell* 68 635-645 (1992).
9. Kemble, G. W., Bodian, D. L., Rose', J., Wilson, I. A. and White, J. M., "Intermonomer disulfide bonds impair the fusion activity of the influenza hemagglutinin." *J Virol* in press (1992).

10. Hirsch, M. S. and Kaplan, J. C., "Antiviral Agents." In B. N. Fields and D. M. Knipe (Eds.), *Virology*. (Raven Press, Ltd, New York, 1990) pp. 441-468.
11. Presber, H. W., Schroeder, C., Hegenscheid, B., Heider, H., Reefschlager, J. and Rosenthal, H. A., "Antiviral activity of norakin (triperiden) and related anticholinergic antiparkinsonism drugs." *Acta virol* 28 501-507 (1984).
12. Ghendon, Y., Markushin, S., Heider, H., Melnikiv, S. and Lotte, V., "Hemagglutinin of influenza A virus is a target for the antiviral effect of norakin." *J Gen Virol* 67 1115-1122 (1986).
13. Stegmann, T., Booy, F. P. and Wilschut, J., "Effects of low pH on influenza virus. Activation and inactivation of the membrane fusion capacity of the hemagglutinin." *JBC* 262 17744-17749 (1987).
14. Sato, Y. B., Kawasaki, K. and Ohnishi, S.-I., "Hemolytic activity of influenza virus hemagglutinin glycoproteins activated in mildly acidic environments." *PNAS* 80 3153-3157 (1983).
15. Belshe, R. B. and Hay, A. J., "Drug resistance and mechanisms of action on influenza A viruses." *J Respir Disease* 10 S52-S61 (1989).
16. Sugrue, R. J., Bahadur, G., Zambon, M. C., Hall-Smith, M., Douglas, A. R. and Hay, A. J., "Specific structural alteration of the influenza hemagglutinin by amantadine." *EMBO* 9 3469-3476 (1990).
17. Lubran, M. M., "Bacterial toxins." *Annals Clin Lab Sci* 18 58-71 (1988).

18. Sprang, S. R., Acharya, K. R., Goldsmith, E. J., Stuart, D. I., Varvill, K., Fletterick, R. J., Madsen, N. B. and Johnson, L. N., "Structural changes in glycogen phosphorylase induced by phosphorylation." *Nature* 336 215-221 (1988).
19. Fox, M. P., Otto, M. J. and McKinlay, M. A., "Prevention of rhinovirus and poliovirus uncoating by WIN 51711, a new antiviral drug." *Antimicrob Agents Chemo* 30 110-116 (1986).
20. Gruenberger, M., Pevear, D., Diana, G. D., Kuechler, E. and Blaas, D., "Stabilization of human rhinovirus serotype 2 against pH-induced conformational change by antiviral compounds." *J Gen Virol* 72 431-433 (1991).
21. McSharry, J. J., Caliguiri, L. A. and Eggers, H. J., "Inhibition of uncoating of poliovirus by arildone, a new antiviral drug." *Virology* 97 307-315 (1979).
22. Icenogle, J., Shiwen, H., Duke, G., Gilbert, S., Rueckert, R. and Anderegg, J., "Neutralization of poliovirus by a monoclonal antibody: kinetics and stoichiometry." *Virology* 127 412-425 (1983).
23. Beddell, C. R., Goodford, P. J., Norrington, F. E., Wilkinson, S. and Wootton, R., "Compounds designed to fit a site of known structure in human hemoglobin." *Br J Pharmac* 57 201-209 (1976).
24. Beddell, C. R., Goodford, P. J., Kneen, G., White, R. D., Wilkinson, S. and Wootton, R., "Substituted benzaldehydes designed to increase the oxygen affinity of human hemoglobin and inhibit the sickling of sickle erythrocytes." *Br J Pharmac* 82 397-407 (1984).

25. Rott, R., Orlich, M., Klenk, H.-D., Wang, M. L., Skehel, J. J. and Wiley, D. C., "Studies on the adaptation of influenza viruses to MDCK cells." *EMBO* 3 3329-3332 (1984).
26. Daniels, R. S., Downie, J. C., Hay, A. J., Knossow, M., Skehel, J. J., Wang, M. L. and Wiley, D. C., "Fusion mutants of the influenza virus hemagglutinin glycoprotein." *Cell* 40 431-439 (1985).
27. Korolkovas, A. and Burckhalter, J. H., "Development of Drugs." In Essentials of Medicinal Chemistry. (John Wiley & Sons, New York, 1976) pp. 12-43.
28. DesJarlais, R. L., Seibel, G. L., Kuntz, I. D., Furth, P. S., Alvarez, J. C., Ortiz de Montellano, P. R., DeCamp, D. L., Babe', L. M., et al., "Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease." *PNAS* 87 6644-6648 (1990).
29. Rutenber, E., manuscript in preparation.
30. Shoichet, B. K., Molecular Docking: Theory and Application to Recognition and Inhibitor Design. University of California, 1991.
31. Shoichet, B. K., personal communication.
32. Weis, W. I., Brunger, A. T., Skehel, J. J. and Wiley, D. C., "Refinement of the influenza virus hemagglutinin by simulated annealing." *JMB* 212 737-761 (1990).
33. Shoichet, B. K., Bodian, D. L. and Kuntz, I. D., "Molecular docking using shape descriptors." *J Comp Chem* 13 380-397 (1992).
34. Pearlman, R. S., "3D searching: an overview of a new technique for computer-assisted molecular design." *NIDA Research Monograph* 112 62-77 (1991).
35. Dixon, J. S., personal communication.

36. Leach, A. R. and Kuntz, I. D., "The conformational analysis of flexible ligands in macromolecular receptor sites." *J Comp Chem* in press.
37. DesJarlais, R. L., Sheridan, R. P., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R., "Docking flexible ligands to macromolecular receptors by molecular shape." *J Med Chem* **29** 2149-2153 (1986).
38. Lewis, R. A., Roe, D. C., Huang, C., Ferrin, T. E., Langridge, R. and Kuntz, I. D., "Automated site-directed drug design using molecular lattices." *J Mol Graph* in press.
39. Blaney, J., personal communication.
40. Seibel, G. L., personal communication.
41. DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R., "Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure." *J Med Chem* **31** 722-729 (1988).
42. Meng, E. C., Shoichet, B. K. and Kuntz, I. D., "Automated docking with grid-based energy evaluation." *J Comp Chem* **13** 505-524 (1992).
43. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr and Weiner, P., "A new force field for molecular mechanical simulation of nucleic acids and proteins." *JACS* **106** 765-784 (1984).

Chapter 9

Further Studies on the Conformational Change of Influenza Hemagglutinin

Introduction

Membrane fusion is an early and essential step mediating the entry of enveloped viruses into susceptible cells [1, 2]. All enveloped viruses possess a specific protein that mediates this fusion event [1, 2]. Of these integral membrane glycoproteins the influenza hemagglutinin (HA)¹ is best characterized [3]. Each of the three identical monomers comprising the HA trimer is composed of two disulfide-bonded polypeptide chains, HA1 and HA2. The extracellular portion of this membrane protein has been described as having three domains: a globular head domain possessing receptor binding activity and major antigenic determinants, a hinge region, and, protruding from the membrane, a stem region where a sequence critical for fusion, the fusion peptide, is located [1, 2, 3, 4]. In the normal replication cycle, the virion binds to receptors on the cell surface and is endocytosed. The low endocytic pH triggers a conformational change in HA that releases the fusion peptides from

¹Abbreviations used: HA, hemagglutinin; BHA, the soluble ectodomain of hemagglutinin; wt-HA, wild type HA produced by the Wtm8005 cell line; Cys-HA, site directed mutant with substitutions T212C, N216C in HA1; cpm, counts per minute; RT, room temperature.

the trimer interface. The viral and endocytic membranes then fuse, releasing the viral ribonucleoprotein core into the cell cytoplasm [1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14].

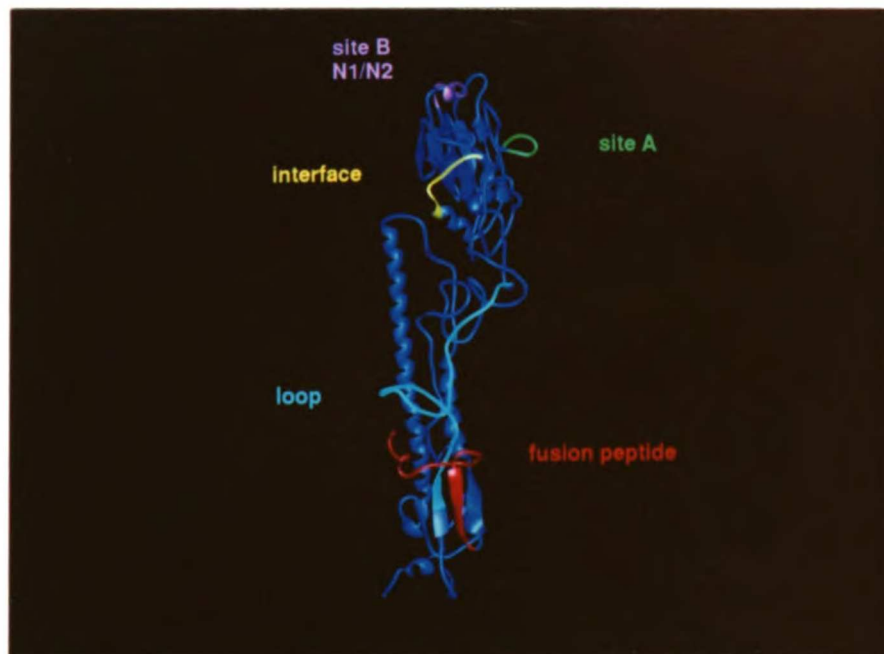


Figure 9-1. Ribbon diagram of the HA monomer ectodomain showing locations of antibody epitopes. Site A is shown in green; site B (N1/N2), magenta; fusion peptide, red; yellow, interface; cyan, loop.

A panel of anti-peptide antibodies raised against sequences located throughout the trimer (Figure 9-1) has been used to study the detailed mechanism of the fusion-inducing rearrangements [13]. Results from these experiments define a two-stage model for the conformational change (Figure 9-3). In the first stage, occurring within 60 seconds at 37°C, changes occur in the stem region of the

trimer, the most important being exposure of the fusion peptides. In a slower step, with a half-time of approximately 4 minutes at 37°C, the globular heads dissociate substantially from one another. Characterization of the X31 HA at 0°C and the HA from the Japan strain of influenza at 37°C has supported this two-stage conformational change model [15, 16, 17]. While prior work has demonstrated the importance of fusion peptide release for fusion [18, 19] it has remained unclear whether the rearrangements in the head domains are required for fusion or whether they represent a post-fusion event [13, 15, 16, 20].

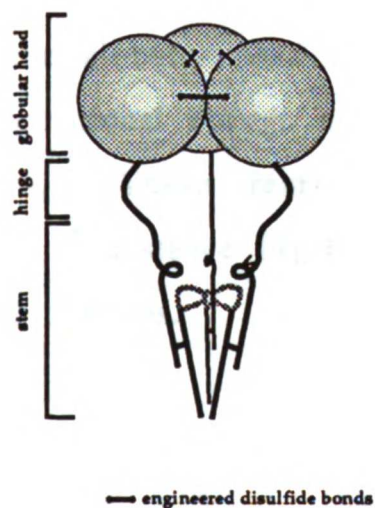


Figure 9-2. Representation of the HA trimer showing the three subdomains of the HA ectodomain (globular head, hinge, stem). The fusion peptides are drawn as stippled lines. The locations of the disulfide bonds engineered into Cys-HA at residues HA1 212 and 216 are also shown. The mutant was made and analyzed in the lab of Dr. Don Wiley [21].

The role of dissociation of the head domains was explored in two systems in which this separation is inhibited. It has been reported that at low temperature and low pH, HA does not undergo the second stage changes but the fusion peptides are released and the protein becomes fusogenic [16, 17]. This phenotype is different from that of a mutant HA, Cys-HA, in which the heads of adjacent monomers are covalently joined via engineered disulfide bonds (Figure 9-2). When exposed to acidic conditions, Cys-HA does not change conformation nor does it induce syncytia formation or red blood cell lysis [21]. The conformational changes of native hemagglutinin at 0°C and Cys-HA at 37°C were examined in more detail with the panel of conformation-specific antibodies (Figure 9-1). These studies revealed that a change, most likely representing partial dissociation of the globular head domains, occurs during the first stage of the conformational change concurrent with the changes occurring in the stem. These rearrangements are apparently cooperative over the ~ 75 Å distance [4] between the fusion peptides and the tips of the head domains.

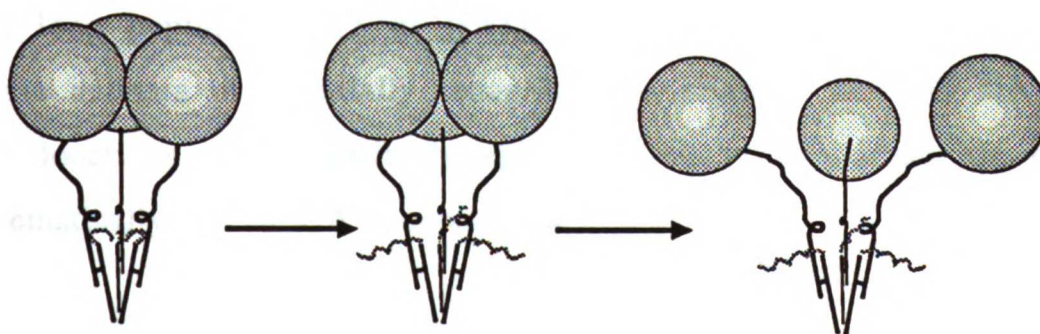


Figure 9-3. Schematic representation of the proposed two stage mechanism of the conformational change of influenza hemagglutinin

[13]. Exposure to low pH first triggers early events including exposure of the fusion peptides (gray wavy lines). In a second, slower step, the globular head domains (shaded circles) dissociate substantially from one another. Figure drawn by Jason Rosé.

Materials and Methods

Antibodies

The site A mouse monoclonal antibody was the gift of Dr. John Skehel. It reacts with properly folded neutral and low pH treated HA [4, 22]. The N1 and N2 mouse monoclonal antibodies, gifts of Dr. Ari Helenius, are specific for the neutral HA trimer [16, 23]. These antibodies react with site B of the HA molecule at the tip of the globular head [4, 16]. The interface mouse monoclonal antibody (H26D08), a gift of Dr. Ian Wilson, reacts with a peptide spanning amino acids 98-106 of the X31 HA1 sequence. This epitope is completely buried within the trimer interface and is inaccessible to antibody molecules at neutral pH [13, 24]. However, the antibody does react with HA trimers that have been exposed to low pH [13]. Rabbit antisera against the fusion peptide (residues 1-29 of HA2) and loop peptide (14-52 of HA1) were gifts of Dr. Richard Lerner. These antisera preferentially recognize the acidified form of HA [13]. The locations of the antibody epitopes used in this study are shown schematically on an HA monomer in Figure 9-1.

Cells and Reagents

Cells expressing wild type (wt) X31 HA (H3 subtype, Wtm8005 cell line) and Cys-HA (X31 HA with the substitutions T212->C and N216->C) were the generous gift of Dr. Don Wiley [21]. Cys-HA and Wtm8005 cells were maintained in G418 media (minimal essential media-alpha without nucleosides, 10% supplemented bovine calf serum (SCS), 2 mM glutamine, 100 U/ml penicillin, 100 mg/ml streptomycin, 600 μ g/ml geneticin (Gibco BRL, Gaithersburg MD), 0.3 μ M methotrexate). CV-1 cells were maintained in CV-1 growth media (DME, 10% SCS, 2 mM glutamine, 100 U/ml penicillin, 100 mg/ml streptomycin). Unless otherwise indicated all reagents for the maintenance of tissue culture cells were obtained from the UCSF Cell Culture Facility (San Francisco, CA) and all biochemical reagents were purchased from Sigma (St. Louis, MO).

Purification of HA from cell lines

Five T-175 flasks of cells expressing either wt-HA or Cys-HA were metabolically labeled with 0.5 mCi/flask of [³⁵S] Express Label (NEN, Boston, MA) for 18 hours at 37°C in MEM Select-Amine lacking cysteine and methionine (Gibco BRL, Gaithersburg, MD). Unless otherwise indicated, the cells were incubated with 10 μ g/ml TPCK-trypsin in RPMI-1640 for 8 minutes at room temperature (RT) to convert the cell HA precursor to HA1 and HA2. The cleavage reaction was quenched by the addition of 50 μ g/ml STI (soy trypsin inhibitor) in RPMI-1640. The cells were removed from the dish by

scraping into PBS, collected by centrifugation at $115 \times g$ for 5 minutes, and lysed for 30 minutes on ice in 1% nonidet-P40 (NP40)/0.1 M Tris pH 7.5 containing a protease inhibitor cocktail (1 mM PMSF (phenyl methyl sulfonyl fluoride), 1 $\mu\text{g/ml}$ pepstatin A, 2 $\mu\text{g/ml}$ leupeptin, 4 $\mu\text{g/ml}$ aprotinin, 10 $\mu\text{g/ml}$ antipain, 50 $\mu\text{g/ml}$ benzamidine, 10 $\mu\text{g/ml}$ STI, 0.5 mM IAA (iodoacetamide)). Following centrifugation in a TLA100.3 rotor (Beckman, Fullerton, CA) at $70,000 \times g$ for 30 minutes at 4°C , the detergent soluble HA was purified by ricin affinity chromatography and sucrose gradient purification as described [13]. Fractions corresponding to 9S trimers were pooled and used for immunoprecipitation analyses. The preparations were approximately 80% pure as determined by laser densitometry, the major contaminant being a 116 kD polypeptide that does not crossreact with anti-HA antibodies (data not shown).

Purification of BHA from infected cells

[^3H]Leu-BHA was prepared similarly to [^{35}S]-HA [13]. T-150 flasks of 60-70% confluent CV-1 cells were infected with 6 ml of 10,000 HAU/ml inoculum in DME without serum. Cells were incubated for 90 minutes at 37°C , 5% CO_2 with periodic agitation. After addition of CV-1 growth media, incubation proceeded for a further 4 hours. The cells were then washed with leu- media (MEM (GIBCO), 2.2 g/l NaHCO_3 , 58 mg/l lysine, 15 mg/l methionine, pH 7.3) and incubated for 90 minutes at 37°C , 5% CO_2 . Labeling continued overnight in leu-media with 2.5% SCS and 1 mCi/flask [^3H]-leucine. Cells were washed and collected in PBS, pelleted for 5 minutes at 170

x g at RT, then lysed on ice for 15 minutes in 3.5 mls cell lysis buffer (10 mM MES (2-[N-Morpholino]ethanesulfonic acid), 10 mM HEPES, 100 mM NaCl, 0.1% NP40, pH 7.1) containing 1 mM PMSF and 10 µg/ml aprotinin. Cell debris was pelleted by centrifugation in a TLA 100 for 30 minutes at 70,000 x g. The supernatant was digested for 16 hours at RT with 0.1 mg/ml bromelain (Sigma), 20 mM β-mercaptoethanol (Biorad). The reaction was quenched by addition of 1 mM N-ethylmaleimide (Sigma) and the solution immediately applied to a ricin-agarose column. After washing for approximately 8 hours at 4°C with PBS/0.1% NP40 the protein was eluted with 0.2 M galactose/PBS/0.1% NP40. Peak fractions were pooled, concentrated, and purified on a 5%-25% w/vol sucrose/MES Saline (20 mM MES, .13 M NaCl, pH 7)/0.1% NP40 gradient in an SW41 rotor for 16 hours at 170,000 x g and 4°C. The fractions corresponding to the 9S trimer peak were collected and pooled.

Immunoprecipitation

Samples of purified [³⁵S]wt- or [³⁵S]Cys-HA in MS buffer (0.13 M NaCl, 10 mM MES, pH 7.0) containing 0.1% NP40 were treated at pH 5.0 for the indicated times at 37°C, reneutralized, and reacted with primary antibody for 2 hours at RT. Final dilutions of antibodies were as follows: fusion peptide and loop 1:5, N1 and interface 1:100, N2 1:25, and site A 2.9 µg/ml. Protein-antibody complexes were precipitated for 1 hour with 20 µl of protein A agarose (Schleicher and Schuell, Keene, NH), washed twice with 0.5 M

NaCl, 10 mM Tris, pH 8.0, resuspended in 10% sodium dodecyl sulfate, and processed for scintillation counting.

Scintillation Proximity Assay

[³H]Leu-BHA was diluted to 10,000 cpm/100 µl in MSSH buffer (10 mM HEPES, 10 mM MES, 10 mM succinate, 0.10 M NaCl, pH 7.0) containing 0.1% NP40. Solutions were cooled to 0°C for 30 mins on ice or warmed to 37°C for 15 minutes. A predetermined amount of 1N acetic acid was added to bring the reactions to pH 5.0. Following incubation at the given temperature for the appropriate time, reactions were reneutralized with 1N NaOH. 100 µl of the protein solution was added to scintillation vials containing 100 µl protein A-SPA beads (Amersham), 100 µl MSSH buffer, 5% fetal bovine serum, and antibody. Final dilutions of antibodies were as follows: fusion peptide and N2 1:300, loop 1:150, N1 1:3000, interface 1:600, and site A 0.5 µg/ml. Reactions were incubated overnight at RT with constant shaking. Cpm were detected in a Beckman LS3801 scintillation counter without addition of liquid scintillant.

Computer Modeling of HA

The ribbon drawing of HA was produced using the MidasPlus software system from the Computer Graphics Laboratory, University of California, San Francisco [25] and displayed on an Iris 4D/80GT workstation (Silicon Graphics, Mountain View, CA). Atomic

coordinates are from entry 2hmg [26] in the Brookhaven Protein Databank [27, 28].

Results

Purified, detergent solubilized [^{35}S]-labeled wt-HA and Cys-HA trimers were compared for their ability to react with a set of conformation-specific antibodies both before and after exposure to low pH. The epitopes recognized by these probes are displayed on a monomer of HA in Figure 9-1. As shown previously [13, 16, 29] and in Table 9-1, exposure of wt-HA to low pH resulted in its conversion to the low pH form as evidenced by its acquisition of antibody

EPITOPE	WT-HA			CYS-HA		
	pH7	pH5	Δ	pH7	pH5	Δ
FUSION PEPTIDE	1	36	+35	1	10	+9
LOOP	1	43	+42	1	7	+6
N1	44	1	-43	37	25	-12
INTERFACE	0	10	+10	0	1	+1

Table 9-1. Comparison of antibody binding to wt- and Cys-HA. [^{35}S] wt- and Cys-HA trimers were purified from cells and treated at the indicated pH for 5 (fusion peptide, loop, N1) or 15 (interface) minutes at 37°C, reneutralized, and precipitated with the indicated antibody. The percentage of input counts precipitated is shown. Each entry is the average of two samples. This experiment was performed three separate times with similar results. Δ represents the difference between the %cpm precipitated between pH5 and pH7 treated wt- or Cys-HA.

reactivity to fusion peptide, loop, and interface epitopes as well as its loss of reactivity with N1, a neutral pH, trimer specific antibody [23]. The conformational change as detected by binding of these antibodies was more modest in Cys-HA. Overall, this analysis with four conformation specific antibodies against defined epitopes indicates that Cys-HA is significantly impaired in its ability to undergo the low pH-induced conformational changes associated with wt-HA, even those that occur at a great distance in the stem of the molecule.

To extend this analysis, we compared wt- and Cys-HA for the time course and extent of acquisition of reactivity to the anti-fusion peptide antibody and loss of reactivity to N2, another trimer specific antibody. Low pH-treated wt-HA quickly acquired reactivity with the fusion peptide antiserum and concomitantly lost its reactivity with N2 (Figure 9-4). In contrast, although the time courses were similar, the extent of these conformational changes in Cys-HA was greatly reduced (Figure 9-4). Nevertheless, for both wt- and Cys-HA good correlation was observed between the time courses and extents of fusion peptide exposure and loss of the neutral-specific epitope, N2. These results suggest that loss of the N2 epitope located at the tip of the globular head domain interface occurs concomitantly with exposure of the fusion peptide epitope. When the globular head domains are physically joined, both of these changes are markedly and commensurately inhibited.

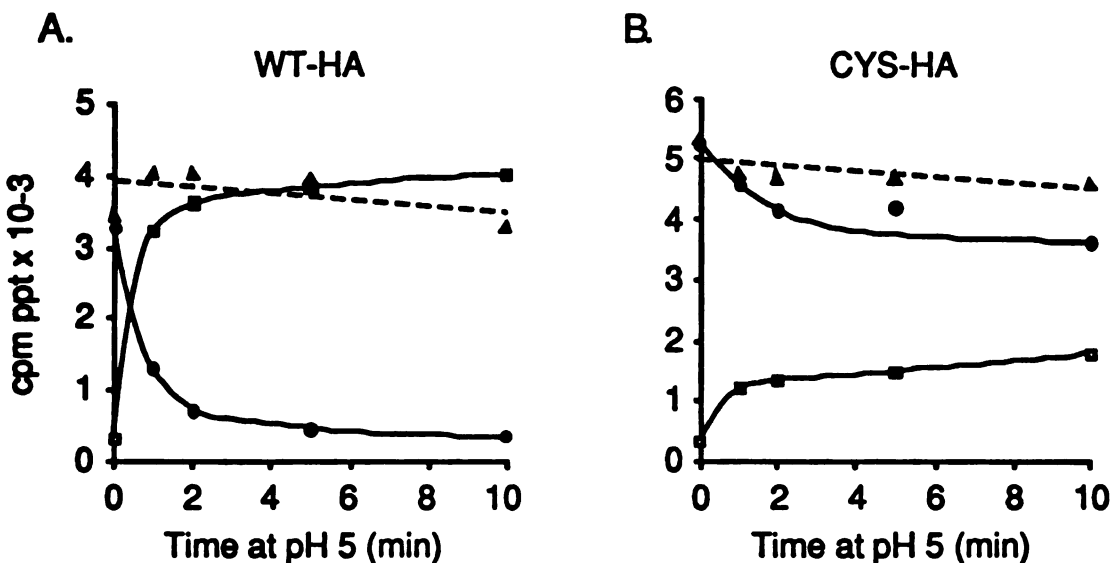


Figure 9-4. Low pH-induced conformational changes in wt- and Cys-HA. [³⁵S]wt-HA (A) or [³⁵S]Cys-HA (B) were incubated at pH 5.0 for the indicated times then reneutralized. The proteins were incubated with either the site A (▲), N2 (●), or fusion peptide (□) antibody for 2 hours, the protein-antibody complexes were precipitated with protein A agarose, washed, and processed for scintillation counting as described. The number of cpm precipitated under these conditions is shown.

The cooperativity of the conformational changes was further explored by comparing the behavior of wild type hemagglutinin at 0°C to that at 37°C. The assay required used of the solubilized ectodomain of X31 HA (BHA). BHA has proved to be a reliable model for intact HA in assays not requiring membrane attachment [6, 13]. Previous studies have shown that at low temperatures major dissociation of the head domains is inhibited but exposure of the fusion peptides still occurs [16, 17]. Consistent with results from intact viruses [16], the fusion peptides were released at a similar rate and to the same extent at 0° and 37° (Table 9-2 and Figure 9-5) However, unlike earlier results, the N1 epitope was lost at 0° (Table

9-2 and Figure 9-5). This change occurred at the same rate and to the same extent as at 37° (Figure 9-5). As with wt-HA and Cys-HA, these changes in the head domains occurred within 1 minute, concurrently with exposure of the fusion peptides.

EPITOPE	pH 7	pH 5	
		37°C	0°C
FUSION PEPTIDE	0.3	15.4	15.0
N1	16.1	2.1	4.3
INTERFACE	0.0	1.6	0.1
A	17.1	10.9	12.4

Table 9-2. Comparison of antibody binding to X31 BHA at 0° and 37°C. [³H]Leu-BHA trimers were purified from infected cells and treated at the indicated pH for 5 (37°) or 15 (0°) minutes, reneutralized, and precipitated with the indicated antibody by scintillation proximity assay. The percentage of input counts precipitated is shown. Each entry is the average of 1-6 samples from independent experiments. The cpm precipitated by the interface antibody are too low to draw significant conclusions.

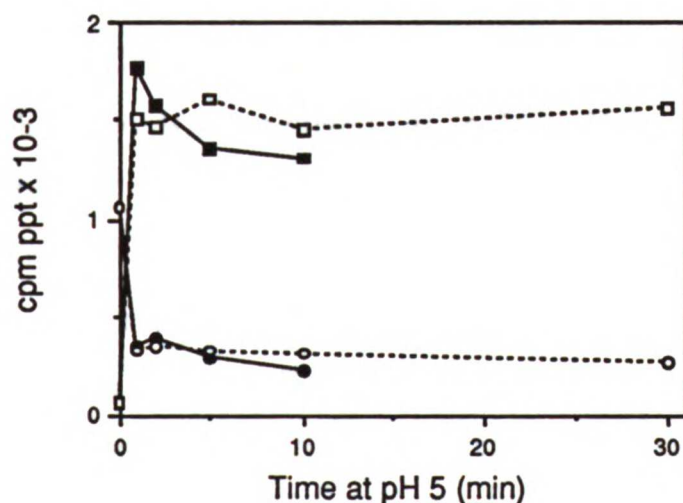


Figure 9-5. Low pH-induced conformational changes in X31 BHA at 0° and 37°C. [³H]Leu-X31 BHA was incubated at pH 5.0 for the

indicated times at 37°C (solid lines) or 0°C (dotted lines) then reneutralized. Radioactive counts were determined by scintillation proximity assay with either the fusion peptide (squares) or N1 (circles) antibody. The number of cpm precipitated under these conditions is shown. Data is from a single experiment.

Fusion Activity of Cys-HA

The ability of cell surface Cys-HA to change conformation and to mediate fusion was analyzed by Dr. George Kemble [30]. One characteristic of the low pH conformation of hemagglutinin is ability to be degraded by proteinase K. This property is acquired during the first stage of the conformational change [15]. Only 25% of cell surface Cys-HA became susceptible to proteinase K digestion following low pH treatment, whereas nearly all of wt-HA was degraded. Treatment of Cys-HA with dithiothreitol (DTT) increased the proportion of cell surface Cys-HA that became susceptible to proteinase K digestion.

Cys-HA is significantly impaired in its ability to mediate fusion. Fluorescence dequenching experiments revealed that fusion mediated by Cys-HA occurred at only ~15% the rate of wt-HA mediated fusion, and only after a substantial lag. The decreased activity was not due to differential expression since both Cys-HA and wt-HA were present in equivalent amounts in the surfaces of these cells. Reduction with DTT eliminated the lag and restored the rate of fusion to that observed with wt-HA expressing cells treated with the same amount of DTT.

SDS-PAGE analysis revealed that 90% of the cell surface Cys-HA migrated as trimer, and that reduction with DTT both decreased the proportion of trimer and increased the amount of monomer. This analysis suggests that the engineered disulfide bonds inhibit Cys-HA from changing conformation and from mediating fusion. The low level of activity observed in the fusion assay and in the conformational change assays with cell surface or purified Cys-HA is attributable to incomplete formation of the engineered disulfide bonds.

Discussion

Membrane fusion is important for delivering the genome of an enveloped virus into a susceptible cell as well as for many aspects of a cell's normal function [1, 2, 31, 32]. Upon exposure to low pH the HA molecule retains its overall secondary structure, but undergoes rearrangements that are a prerequisite for fusion. The fusion peptides are extruded from within the trimer stem and the heads separate substantially from one another exposing epitopes within the interface of the globular head domains [1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. While the changes in the stem region are clearly critical for fusion, the role of rearrangements in the head domains has remained unclear.

Partial dissociation of the globular head domains occurs during the first stage of the low pH induced conformational change. Previous work using conformation-specific anti-peptide antibodies demonstrated that HA changes

conformation in two stages [13]. The first stage is defined as those changes that occur in less than 60 seconds at 37°C. Marked changes occur in the stem region of the HA during stage 1 including exposure of the loop, C-terminus of HA1, and fusion peptide epitopes. Changes during the second stage ($t_{1/2}$ ~4 minutes) are characterized by exposure of peptide sequences located between the head groups. Biophysical measurements and electron microscopic observations suggest that during the second stage of the conformational change the head groups separate substantially from one another [1, 10, 13].

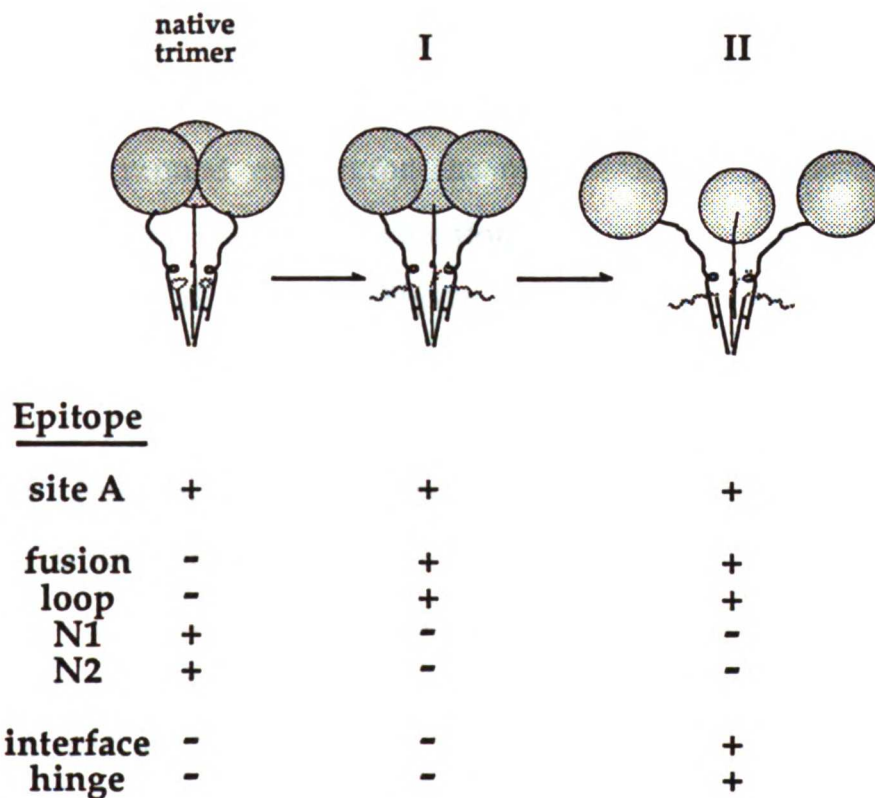


Figure 9-6. Depiction of the current model for low pH-induced conformational changes in HA. In stage I, the globular heads have partially dissociated from each other and epitopes in the stem, the fusion peptide (dashed line/squiggle) and loop peptide, have been

exposed. In stage II, major dissociation of the globular head domains has occurred. The lower part of the figure indicates the presence (+) or absence (-) of a particular epitope. Figure drawn by Jason Rosé.

A change in the globular head domains, as evidenced by the loss of the N1 and N2 epitopes, occurs during the first stage of the conformational change concomitant with fusion peptide exposure (Figures 9-4 and 9-5). The extents of the changes are commensurately diminished in Cys-HA (Table 9-1 and Figure 9-4). Since the N1 and N2 epitopes appear coincident with trimerization of the molecule [23] and since there are no gross secondary structural changes to HA after low pH treatment [3, 10, 33], this change likely represents partial dissociation of the head domains. To account for this finding a refined model for the conformational change in HA is proposed in which partial dissociation of the globular head domains occurs during the first stage (Figure 9-6). As in the original model, major dissociation occurs during the second stage.

Role of Globular Head Domain Dissociation in HA-Mediated Fusion. Two recent studies support the notion that *major* dissociation of the globular head domains, operationally defined here as maximal separation of the globular head domains from one another, is not required for fusion. With X31 HA (an H3 subtype) at 0°C or Japan HA (an H2 subtype) at 37°C the second stage of the conformational change appears to be inhibited, yet fusion proceeds [15, 16, 17]. Also, the time course of fusion of X31 virus [20] is faster than that of the second stage changes [13]. These findings do not, however, rule out the possibility that partial

separation of the globular head domains is required for fusion, at least for the X31 strain. Since inhibition of both major and minor dissociation is inhibited in Cys-HA and since major dissociation is apparently not required for fusion, partial dissociation of the globular head domains seems to be necessary to promote fusion peptide release.

In addition to their relevance to the fusion mechanism of the influenza HA, the results speak to the design of antiviral agents. In theory, inhibition of fusion peptide exposure and thus fusion should prevent viral replication. Since locking the heads together such that the N1/N2 epitopes are not lost inhibits the rearrangements in the stem, the conformational changes are apparently cooperative throughout much of the trimer. The residual ability of Cys-HA to change conformation is presumably due to incomplete formation of the engineered disulfide bonds [30]. Therefore, inhibitors can be designed to interact directly with the stem region, or with sites distant from the fusion peptide. Both strategies could have the same ultimate effect, preventing fusion peptide exposure, and hence the fusion activity of influenza virus.

References

1. Stegmann, T., Doms, R. W. and Helenius, A., "Protein-mediated membrane fusion." *Ann Rev Biophys Biophys Chem* 18 187-211 (1989).

2. White, J. M., "Viral and cellular membrane fusion proteins." *Ann Rev Physiol* 52 675-697 (1990).
3. Wiley, D. C. and Skehel, J. J., "The structure and function of the hemagglutinin membrane glycoprotein of the influenza virus." *Ann Rev Biochem* 56 365-394 (1987).
4. Wilson, I. A., Skehel, J. J. and Wiley, D. C., "Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution." *Nature* 289 366-373 (1981).
5. Daniels, R. S., Downie, J. C., Hay, A. J., Knossow, M., Skehel, J. J., Wang, M. L. and Wiley, D. C., "Fusion mutants of the influenza virus hemagglutinin glycoprotein." *Cell* 40 431-439 (1985).
6. Doms, R. W., Helenius, A. and White, J., "Membrane fusion activity of the influenza virus hemagglutinin: the low pH-induced conformational change." *JBC* 260 2973-2981 (1985).
7. Graves, P. N., Schulman, J. L., Young, J. F. and Palese, P., "Preparation of influenza virus subviral particles lacking the HA1 subunit of hemagglutinin: unmasking of cross-reactive HA2 determinants." *Virology* 126 106-116 (1983).
8. Ruigrok, R. W. H., Martin, S. R., Wharton, S. A., Skehel, J. J., Bayley, P. M. and Wiley, D. C., "Conformational changes in the hemagglutinin of influenza virus which accompany heat-induced fusion of virus with liposomes." *Virology* 155 484-497 (1986).
9. Ruigrok, R. W. H., Aitken, A., Calder, L. J., Martin, S. R., Skehel, J. J., Wharton, S. A., Weis, W. and Wiley, D. C., "Studies on the structure of the influenza virus hemagglutinin at the pH of membrane fusion." *J Gen Virol* 69 2785-2795 (1988).

10. Skehel, J. J., Bayley, P. M., Brown, E. B., Martin, S. R., Waterfield, M. D., White, J. M., Wilson, I. A. and Wiley, D. C., "Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion." *PNAS* 79 968-972 (1982).
11. Webster, R. G., Brown, L. E. and Jackson, D. C., "Changes in the antigenicity of the hemagglutinin molecule of H3 influenza virus at acidic pH." *Virology* 126 587-599 (1983).
12. Wharton, S. A., Ruigrok, R. W. H., Martin, S. R., Skehel, J. J., Bayley, P. M., Weis, W. and Wiley, D. C., "Conformational aspects of the acid-induced fusion mechanism of influenza virus hemagglutinin. Circular dichroism and fluorescence studies." *JBC* 263 4474-4480 (1988).
13. White, J. M. and Wilson, I. A., "Anti-peptide antibodies detect steps in a protein conformational change: low pH activation of the influenza virus hemagglutinin." *J Cell Bio* 105 2887-2896 (1987).
14. Yewdell, J. W., Gerhard, W. and Bachi, T., "Monoclonal anti-hemagglutinin antibodies detect irreversible antigenic alterations that coincide with the acid activation of influenza virus A/PR/834-mediated hemolysis." *J Virol* 48 239-248 (1983).
15. Puri, A., Booy, F. P., Doms, R. W., White, J. M. and Blumenthal, R., "Conformational changes and fusion activity of influenza virus hemagglutinin of the H2 and H3 subtypes: effects of acid pretreatment." *J Virol* 64 3824-3832 (1990).

16. Stegmann, T., White, J. M. and Helenius, A., "Intermediates in influenza induced membrane fusion." *EMBO* 9 4231-4241 (1990).
17. Stegmann, T., Booy, F. P. and Wilschut, J., "Effects of low pH on influenza virus. Activation and inactivation of the membrane fusion capacity of the hemagglutinin." *JBC* 262 17744-17749 (1987).
18. Harter, C., James, P., Bachi, T., Semenza, G. and Brunner, J., "Hydrophobic binding of the ectodomain of influenza hemagglutinin to membranes occurs through the 'fusion peptide'." 264 6459-6464 (1989).
19. Gething, M.-J., Doms, R. W., York, D. and White, J., "Studies on the mechanism of membrane fusion: site-specific mutagenesis of the hemagglutinin of influenza virus." *J Cell Bio* 102 11-23 (1986).
20. Stegmann, T., Hoekstra, D., Scherphof, G. and Wilschut, J., "Fusion activity of influenza virus. A comparison between biological and artificial target membrane vesicles." *JBC* 261 10966-10969 (1986).
21. Godley, L., Pfeifer, J., Steinhauer, D., Ely, B., Shaw, G., Kaufmann, R., Suchanek, E., Pabo, C., et al., "Introduction of intersubunit disulfide bonds in the membrane-distal region of the influenza hemagglutinin abolishes membrane fusion activity." *Cell* 68 635-645 (1992).
22. Daniels, R. S., Douglas, A. R., Skehel, J. J. and Wiley, D. C., "Analyses of the antigenicity of influenza hemagglutinin at the

- pH optimum for virus-mediated membrane fusion." *J Gen Virol* **64** 1657-1662 (1983).
23. Copeland, C. S., Zimmer, K.-P., Wagner, K. R., Healey, G. A., Mellman, I. and Helenius, A., "Folding, trimerization, and transport are sequential events in the biogenesis of influenza virus hemagglutinin." *Cell* **53** 197-209 (1988).
 24. Wilson, I. A., Niman, H. L., Houghton, R. A., Cherenson, A. R., Connolly, M. L. and Lerner, R. A., "The structure of an antigenic determinant in a protein." *Cell* **37** 767-778 (1984).
 25. Ferrin, T. E., Huang, C. C., Jarvis, L. E. and Langridge, R., "The Midas display system." *J Mol Graph* **6** 13-27 (1988).
 26. Weis, W. I., Brunger, A. T., Skehel, J. J. and Wiley, D. C., "Refinement of the influenza virus hemagglutinin by simulated annealing." *JMB* **212** 737-761 (1990).
 27. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., et al., "The Protein Data Bank: A computer-based archival file for macromolecular structures." *J Mol Biol* **112** 535-542 (1977).
 28. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. and Weng, J., In F. H. Allen, G. Bergerhoff and R. Seivers (Eds.), Crystallographic Databases - Information Content. Software Systems. Scientific Applications. (Data Commission of the International Union of Crystallography, Cambridge, 1987) pp. 107-132.
 29. Boulay, F., Doms, R. W., Webster, R. G. and Helenius, A., "Posttranslational oligomerization and cooperative acid

activation of mixed influenza hemagglutinin trimers." *JCB* 106 629-639 (1988).

30. Kemble, G. W., Bodian, D. L., Rose', J., Wilson, I. A. and White, J. M., "Intermonomer disulfide bonds impair the fusion activity of the influenza hemagglutinin." *J Virol* in press (1992).
31. Wilschut, J., "Intracellular membrane fusion." *Curr Op Cell Biol* 1 639-647 (1989).
32. Wilson, D. W., Wilcox, C. A., Flynn, G. C., Chen, E., Kuang, W.-J., Henzel, W. J., Block, M. R., Ullrich, A., et al., "A fusion protein required for vesicle-mediated transport in both mammalian cells and yeast." *Nature* 339 355-359 (1989).
33. Sato, Y. B., Kawasaki, K. and Ohnishi, S.-I., "Hemolytic activity of influenza virus hemagglutinin glycoproteins activated in mildly acidic environments." *PNAS* 80 3153-3157 (1983).

Chapter 10

Differential Docking: Computer-Aided Design of Selective Inhibitors

Introduction

Drugs exert their effects by binding to particular molecules in the body [1]. To date, the development of new drugs targeted against specific receptors has mostly relied on random screening or on deducing chemical features conducive to binding from the structures of known substrates and inhibitors [2]. However, design of novel compounds based on the structures of the receptors themselves is gaining prominence as three-dimensional coordinates of a growing number of pharmacologically interesting receptors become available [3, 4].

Structure-based design methods identify compounds with geometric and chemical properties complementary to those of the intended binding site. Although these methods have led to the development of novel ligands [5, 6, 7, 8], affinity is not by itself sufficient to make an inhibitor a good drug. One of the most important characteristics of a useful drug is specificity. Fortuitous interactions with targets other than the desired receptor can result in serious and undesirable side effects [9]. While the basis of specificity of some inhibitors is unknown, the preference of others for one site

over a highly similar site is attributable to interactions with chemical features differentiating the two sites [10, 11, 12].

Several computer programs are available to assist in the design of novel inhibitors based on steric and/or electrostatic properties of a particular target site [13, 14, 15, 16, 17]. One such program, DOCK, identifies as potential inhibitors small molecules with shapes complementary to that of the desired binding site [18, 19, 20]. This paper describes a new module, DIFFDOCK, that aids in the development of selective inhibitors. Whereas DOCK orients molecules over the entire site, DIFFDOCK identifies only those ligands which could potentially bind to designated regions of the site. The highlighted regions are user-defined and may include those areas which differentiate the site from corresponding sites of homologous receptors.

This paper presents two test cases. In the first case, DIFFDOCK uses the differences between native α -lytic protease and a site-directed mutant to guide the docking of an inhibitor more active against the mutant. In the second case, the differing size of the specificity pockets of two serine proteases, chymotrypsin and elastase, is used to dock a chymotrypsin-specific inhibitor into the active site of that enzyme. In both examples DIFFDOCK is able to reproduce the crystallographically determined configurations of the complexes well within 1 Å rms¹.

¹Abbreviations used: rms, root mean square deviation; PDB, Brookhaven Protein Data Bank; cpu, central processing unit; M192A, α -lytic protease mutant Met 192 -> Ala; LPF, N-acetyl-L-leucyl-L-phenylalanine trifluoromethyl ketone.

Methods

DIFFDOCK is an extension of version 2 of the DOCK program [20]. The DOCK method and parameter definitions are presented elsewhere [18, 19, 20]. The steps of 1) characterization of the negative image of the receptor site, 2) orientation of molecules in the site, and 3) scoring the resulting configurations are summarized here. The DIFFDOCK enhancements required to focus orientations to selected subsites are described in detail below.

Site Characterization

The negative image of a receptor site is described by a set of overlapping spheres of various sizes [18]. The program SPHGEN [18] computes the size and position of the spheres from the accessible molecular surface [21] of the receptor. A single linkage algorithm is used to divide the spheres into nonoverlapping sets called clusters. Each cluster completely fills an invagination on the receptor surface and represents a potential binding site. The centers of the spheres belonging to a single cluster serve as an irregular grid over the site, with each sphere center representing a possible position for a ligand atom. However, only a subset of the atoms will be constrained to the vicinity of a grid point for each orientation of the ligand.

Generation of Orientations

"Matching" is the process of constructing orientations of the small molecule in the receptor site by comparing distances between a subset of the ligand atoms and a subset of the sphere centers. Matching begins by pairing the first ligand atom with the first receptor sphere center. A second atom-sphere center pair is selected such that the distance between the two atoms equals the distance between the two sphere centers within a user-defined tolerance. Additional ligand atom-sphere center pairs are added to the growing match only if the internal distances between all selected atoms approximate the internal distances between the corresponding sphere centers. A minimum of four pairs are required to determine a unique rotation/translation matrix defining an orientation of the ligand in the site [18].

Matching continues until each initial ligand atom has been paired with each receptor sphere. However, the search is not exhaustive as it is not computationally feasible to search all combinations of four spheres and four atoms. Instead, a set of heuristics has been developed to guide the search. The details of the pruning are described fully elsewhere [20].

Typically, tens of thousands of matches are produced for each ligand. The number of orientations generated in a particular region of the site depends on the density of sphere centers characterizing that area. No attempt is made to eliminate redundant configurations.

Differential Docking

Within the limits set by the heuristics, DIFFDOCK computes all possible matches but only selects those that include a user-defined number of flagged spheres. These spheres have been prejudged as representing site regions of particular interest. Matches are checked for flagged spheres after passing the internal distance requirement but before transformation of the ligand coordinates and scoring. Dividing the labeled spheres into two or more sets forces docking to multiple site regions simultaneously. A normal DOCK run is achieved by setting the minimum number of flagged spheres required per match to zero.

Scoring

The coordinates of every atom in the ligand are transformed into the receptor site by applying the rotation/translation matrix defined by each match [18, 22]. Configurations are scored by how well the shape of the ligand in that orientation complements the shape of the site. A regular grid with user-defined resolution is precalculated over the receptor site using a scoring function similar to a van der Waals potential function. If a grid point lies within a user-defined close contact limit of any receptor atom it is considered repulsive and assigned a negative score. A grid point is assigned a score of 1.0 for each receptor atom lying between the close contact distance and a user-set upper limit. Interactions with longer distances are assigned a score of zero. The score for a given

orientation is determined by summing the scores of the grid points lying nearest each ligand atom. In the runs presented here, if any atom of the ligand is assigned a negative score, the entire orientation is discarded. Note that this scoring algorithm was introduced in DOCK2 [20] and provides somewhat different values than earlier approaches [19].

Input

Atomic coordinates for proteins and inhibitors were obtained from the Brookhaven Protein Data Bank [23, 24]. Structures used are 1p02 [25], native α -lytic protease complexed with methoxysuccinyl-alanine-alanine-proline-alanine boronic acid, 1p06 [25], native α -lytic protease complexed with methoxysuccinyl-alanine-alanine-proline-phenylalanine boronic acid, 1p08 [26], α -lytic protease mutant Met 192 \rightarrow Ala complexed with methoxysuccinyl-alanine-alanine-proline-phenylalanine boronic acid, 7gch [27], γ -chymotrypsin complexed with N-acetyl-L-leucyl-L-phenylalanine trifluoromethyl ketone (LPF), and 3est [28], native porcine pancreatic elastase. Elastase was superimposed on chymotrypsin by least squares fitting of the backbone nitrogen and α -carbon atoms of His 57, Asp 102, and Ser 195. The rms of these atoms was 0.12 Å. Distances of elastase atoms to LPF were calculated by superposition of the inhibitor onto the transformed protein coordinates. The PDB structures of the α -lytic enzymes are all in the same coordinate frame.

Molecular surfaces [21] of all proteins were computed at a density of 5.4 points/Å² with DMS, a distributed processing implementation of the MS [29] algorithm (C. Huang, personal communication). A probe radius of 1.4 Å was used and coordinates for the co-crystallized inhibitor, ions and water molecules were not included. Only surfaces of particular regions of interest were used as input to the sphere generation program, SPHGEN. For the α -lytic enzymes, surfaces were calculated from coordinates of the complexed structures 1p06 and 1p08. Spheres were generated from surface contributed by atoms within 15 Å of the centrally located α -carbon of Gly 215A. For chymotrypsin and elastase, the sphere calculation only included surface points derived from atoms within 12 Å of the LPF inhibitor. Spheres for all sites were calculated by running SPHGEN with parameters *surftp* = R, *dentag* = X, *dotlim* = 0.0, *radmax* = 5, *radmin* = 1.3. The largest clusters were used as input to DIFFDOCK without further modification except that for chymotrypsin and elastase, spheres with centers farther than 10 Å from the inhibitor were omitted. The clusters for both chymotrypsin and α -lytic protease encompass subsites S₄ - S₃' [30]².

Spheres characterizing the regions distinguishing chymotrypsin from elastase and the M192A mutant of α -lytic protease from the wild type enzyme were selected graphically. Differential spheres for M192A were required to lie outside the site volume defined by the

²In the terminology proposed by Schechter and Berger [30], P₁ designates the substrate or inhibitor residue on the carbonyl side of the scissile bond and S₁ represents its binding pocket. Residues toward the N terminus of the substrate or inhibitor are identified as P₂, P₃, ..., P_n and bind to subsites S₂, S₃, ..., S_n. Residues to the C terminus are labeled P₁', P₂', ..., P_n' with corresponding binding sites S₁', S₂', ..., S_n'.

cluster of the native enzyme. Spheres for chymotrypsin were selected only if their centers lay in the specificity pocket but outside the volume filled by spheres of the superimposed elastase site.

Differential spheres of the α -lytic protease mutant were grouped into sets based on their distances to the bottom of the site (Figure 10-1a). The protein was reoriented so that the long axis of the specificity pocket was aligned along the y-axis of a standard coordinate system. The bottom of the site was defined as the average y-coordinate (y_{av}) of the two deepest spheres. Distances of all differential spheres to the bottom of the site were calculated as $\sqrt{(y - y_{av})^2}$, where y is the y-coordinate of each sphere center. The 10 spheres with distances between zero and 0.7 Å from the bottom of the site were labeled as set 1. Addition of the 3 spheres with distances between 0.9 and 1.2 Å gave set 2. The remaining differential spheres were at least 1.5 Å from the bottom of the site.

Input to DISTMAP, the program that creates the scoring grid over the receptor site, was: *maxgrid* = 3, *ccon* = 2.4, *polcon* = 2.8, *discut* = 4.5. These parameters are appropriate for scoring noncovalent interactions only. In order to accommodate the close approach of the covalent inhibitors to the hydroxyl oxygen of Ser 195, the scoring grids for all enzymes were computed with that atom omitted.

The DOCK scoring routine was modified slightly for the α -lytic protease runs. The amino terminal residues of the inhibitors bound to the protease are succinylated; however, the methoxysuccinate groups are disordered and no crystallographic coordinates are available. To prevent sterically unreasonable orientations in which

the methoxysuccinate moiety would overlap the protein, a 2 Å radius sphere was attached to the modified amino group for scoring purposes only.

The following values were used for the docking parameters in all runs unless otherwise noted: *dislim* = 1.5, *nodlim* = 4, *ratiom* = 0, *lownod* = 4, *ictbmp* = 0, *expmax* = 0, *lbinsz* = 0.5, *lovlap* = 0.1, *sbinsz* = 1.0, *sovlap* = 0. Assigning *lbinsz* = 1.0, *sovlap* = 0.1 when using the alanyl inhibitor ensured that a comparable number of matches would be generated as with the phenylalanyl inhibitor.

Hardware and Graphics

DMS was run on SUN SPARCstations (SUN Microsystems, Inc., Mountain View, CA). MIDASPLUS [31] was used for visualization of spheres, ligands, proteins, and their surfaces, and for least squares fitting. MIDASPLUS and DIFFDOCK were run on an IRIS 4D/35GT workstation (Silicon Graphics, Inc., Mountain View, CA).

Results

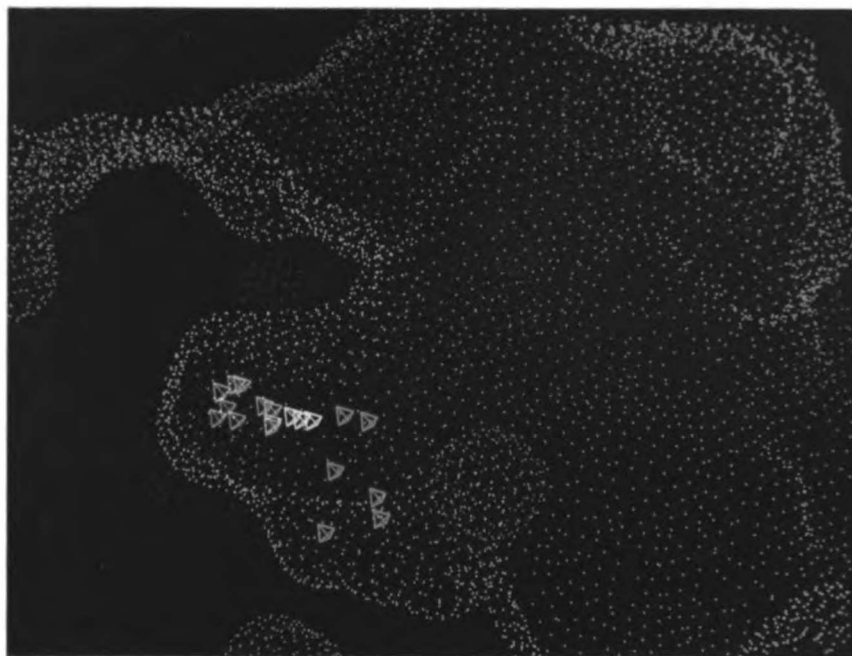
Case 1: α -Lytic Protease

α -Lytic protease is a serine protease that preferentially cleaves peptide bonds on the carboxyl side of small hydrophobic amino acids [32, 33]. Tetrapeptide substrates containing alanine at the P₁ position are hydrolyzed approximately 50,000x more efficiently than the corresponding phenylalanine peptides [25]. Mutation of

methionine 192 to alanine creates an enzyme capable of hydrolyzing a phenylalanine substrate 840,000x more efficiently than the native protease [26]. Both enzymes are inhibited by boronic acid derivatives of these peptides at levels consistent with the observed substrate specificity [26, 34]. Crystal structure analysis of the native and mutant proteins reveals the primary effect of removing the mutation is to enlarge the binding site by 80 Å³ [26]. The prediction that the increased activity of substrates containing phenylalanine in the P₁ position is due to accommodation of the side chain by the expanded pocket has been confirmed by structure determination of the complex of a phenylalanine boronic acid inhibitor with the mutant enzyme [26].

The differences between the structures of the two α-lytic proteins were used to test whether DIFFDOCK can identify the phenylalanine peptide as a selective inhibitor. Molecular surfaces of the inhibitor binding sites in both native and mutant enzymes were computed from the crystal structures of the complexes and used for sphere generation as described in the methods section. Of the 73 spheres characterizing the active site of M192A α-lytic, nineteen were selected as representing the extra site volume arising from the mutation (Figure 10-1a). The phenylalanine inhibitor was then docked into the site. Results of DIFFDOCK runs requiring at least one differential sphere per match are shown in Table 10-1.

(a)



(b)

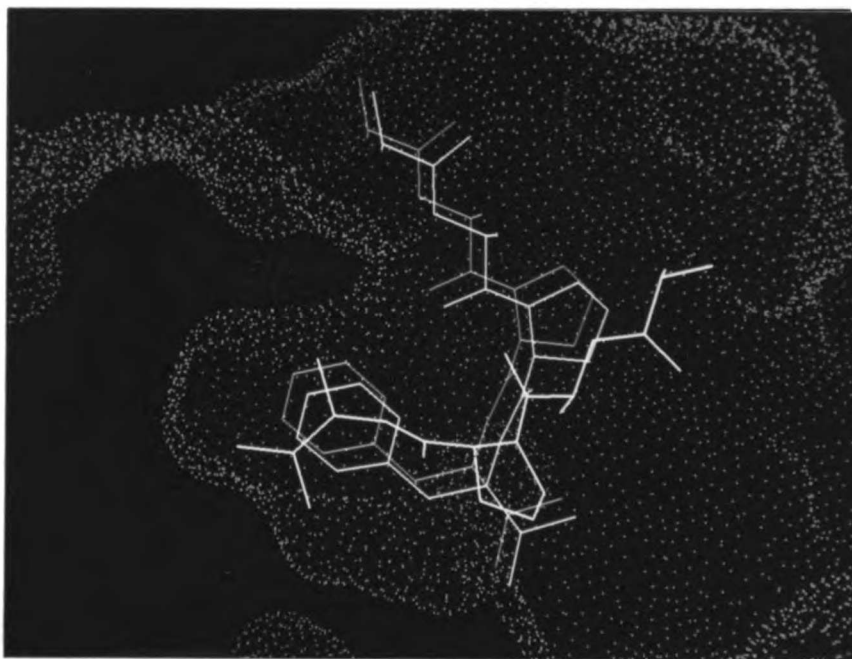


Figure 10-1. Molecular surface of the active site of M192A. a) The nineteen spheres characterizing the extra site volume of the mutant (red, green, dark blue). Tetrahedrons mark the sphere centers. Red

represents the 10 sphere set (see Table 10-2). Addition of the green spheres yields the 13 sphere set. b) Comparison of the crystallographically determined position of the phenylalanine inhibitor (lavendar) with its highest scoring DIFFDOCK orientation (pink). The high scoring docking of the alanyl inhibitor is shown in white.

DIFFDOCK orients the phenylalanine inhibitor into the 1p08 active site within an rms of 0.65 Å from its crystallographically determined coordinates (Figure 10-1b). Limiting the matches to those using at least one of the 19 differential spheres reduces the number of orientations to be transformed and scored by 26%, thereby decreasing computation time by 22%. The savings does not lessen the accuracy of the result as both the lowest rms and the highest scoring orientations found in a full search were recovered. The lowest rms orientation did not receive the highest score in either run due to the highly simplified scoring scheme and the lack of energy minimization.

DIFFDOCK mode	inhibitor	# of matches	high score	rms of high score	best rms	time (sec)
full	Phe	117980	197	0.65	0.42	268.8
differential	Phe	87340	197	0.65	0.42	210.0
full	Ala	98412	113	1.18	1.18	260.6
differential	Ala	80540	108	5.95	3.32	211.8

Table 10-1. DIFFDOCK run on α -lytic mutant M192A. A full run with no flagged spheres is compared to differential docking with 19 flagged spheres representing site volume introduced by the mutation. Coordinates of the inhibitor methoxysuccinyl-AlaAlaProPhe-boronic acid (Phe) are from PDB entry 1p08. Results

from docking the nonselective inhibitor methoxysuccinyl-AlaAlaProAla-boronic acid (Ala) from 1p02 are presented for comparison. Parameters values are listed in the methods. The root mean square deviation of a docked position of the inhibitor from its crystallographic coordinates (rms) was calculated for all orientations receiving a positive score. The rms of the highest scoring orientation as well as the calculated orientation closest to the experimental result are listed. Run time is given as total cpu time in seconds.

The results with the phenylalanine inhibitor show that DIFFDOCK can correctly recognize a selective inhibitor by emphasizing the differential pocket. Using the same requirement, DIFFDOCK should also be able to eliminate nonselective ligands. An alanyl boronic acid peptide inhibits both wild type and mutant proteases [26, 34]. In contrast to the phenyl group of phenylalanine, the short alanine side chain binds to the specificity pocket in a region common to both the native and mutant enzymes. The alanyl inhibitor from 1p02 was docked into the M192A protein in both a full run and a run with the nineteen differential spheres flagged. The docked configurations were then compared to the orientation of the inhibitor resulting from superposition of the complexed wild type enzyme onto the structure of the mutant. In a full run in which the inhibitor is allowed to dock anywhere in the site, the inhibitor is positioned within 1 Å rms of the putative binding mode (Table 10-2). In contrast, a DIFFDOCK run selecting only those orientations in the differential region rejects the low rms configurations. None of the positive scoring orientations found is chemically sensible; all configurations position the methylsuccinyl alanine residue

completely in solvent without interactions with the protein and bury the boronic acid group in the hydrophobic pocket (Figure 10-1b).

DIFFDOCK mode	# of flagged spheres	# of matches	mean score	# of positive scoring orientations	rms of high score	time (sec)
full	73	117908	71	364	0.649	269
differential	19	87340	102	126	0.649	210
differential	13	64611	159	34	0.649	174
differential	10	51281	186	11	0.649	151

Table 10-2. Effect of varying the number of differential spheres. A full run docking AlaAlaProPhe-boronic acid to α -lytic mutant M192A is compared to differential runs in which one of 19, 13, or 10 spheres is required in every match. The shrinking sphere sets characterize progressively deeper regions of the specificity pocket (see Figure 10-1a and methods). DIFFDOCK parameters are listed in the methods. All matches with at least one flagged sphere are scored, and the average score of all positive scoring orientations computed. The root mean square deviation (rms) of the highest scoring orientation to the crystallographic configuration is calculated over the 26 inhibitor atoms in the PDB file. Run time is cpu time in seconds.

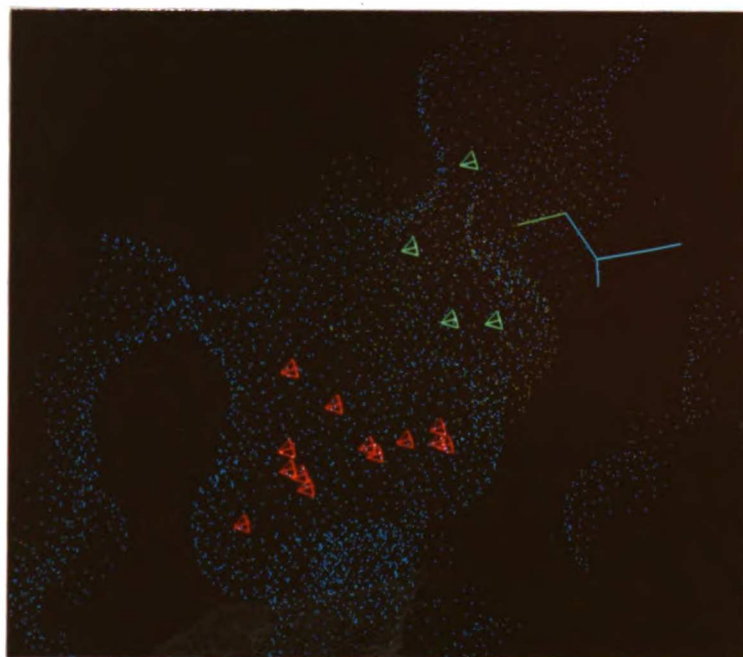
The nineteen flagged spheres span the entire differential region. Table 10-2 shows the effect of gradually removing spheres from the top of the pocket, defined as the region shared between the mutant and the wild type enzymes (Figure 10-1a). Compared to a full run, limiting the search to the bottom of the site reduces the number of acceptable matches by up to 57% without compromising accuracy. As all orientations with scores over 150 lie within 1.0 Å rms of the experimental result (data not shown), the increase in

mean score reveals that tightening the differential criterion focuses the orientations to the crystallographic configuration.

Case 2: Chymotrypsin vs. Elastase

Chymotrypsin and elastase are serine proteases with homologous three-dimensional structures [12] but less than 40% sequence identity [35]. The selectivity of these enzymes is dictated by the structure of their specificity pockets. Chymotrypsin, with a large pocket, hydrolyzes peptide bonds on the carboxyl side of bulky hydrophobic residues [11] whereas elastase, with a significantly shallower site, is thought to prefer peptide substrates with small side chains [12]. Selective inactivation of a single enzyme has been accomplished by synthesis of peptide analogs containing residues preferentially bound by the relevant specificity pocket. One such inhibitor, N-acetyl-L-leucyl-L-phenylalanyl trifluoromethyl ketone (LPF), is a micromolar inhibitor of bovine pancreas chymotrypsin but has a negligible effect on the activity of porcine pancreatic elastase [36]. Superposition of the crystal structure of this elastase onto the structure of γ -chymotrypsin complexed with LPF reveals that the phenylalanine side chain of the inhibitor extends into a region of the specificity pocket unique to chymotrypsin. Therefore, the differences between the S_1 subsites should be able to guide the positioning of the inhibitor into chymotrypsin.

(a)



(b)

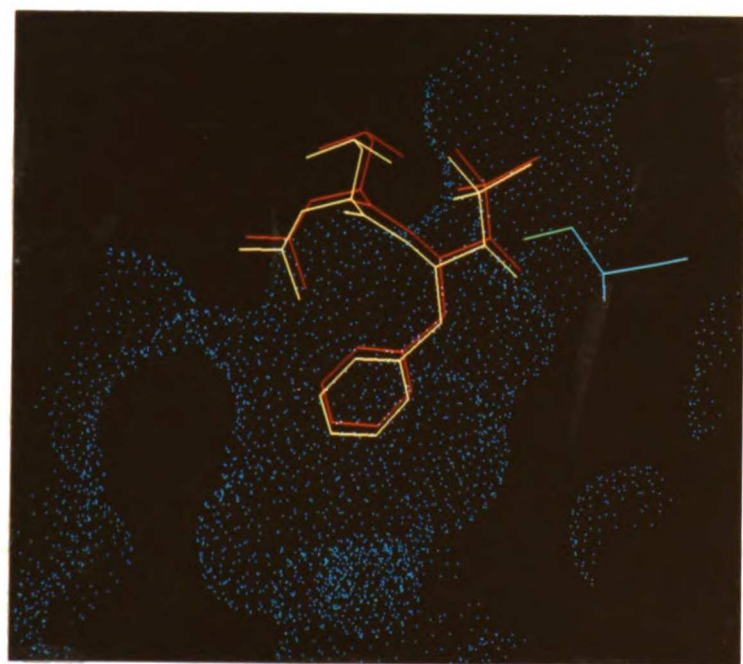


Figure 10-2. Cutaway view of the specificity pocket of γ -chymotrypsin. a) Comparison of the molecular surfaces of the specificity pockets of elastase (yellow) and chymotrypsin (blue). The

proteins were superimposed by least squares fitting as described in the methods. Tetrahedrons locate sphere centers calculated from the chymotrypsin surface. Red spheres represent the region differentiating the specificity pockets of the two enzymes. The hydroxyl group of Ser 195 from chymotrypsin and its 4 closest spheres are shown in green. b) Comparison of the best rms docking (yellow) of LPF to chymotrypsin and its crystallographic configuration (red). The docked orientation received the highest score in all DIFFDOCK runs. Ser 195 of chymotrypsin is displayed in blue with its hydroxyl group colored green.

Molecular surfaces were calculated for the two enzymes, and spheres were generated for chymotrypsin as described in the methods. Of the 83 spheres characterizing the active site, 13 were flagged as representing the differential region (Figure 10-2a). A DIFFDOCK run with LPF targeted to the labeled spheres was compared to a full run with no highlighted spheres. Table 10-3 shows that while the complete docking repositions the inhibitor approximately 0.3 Å from the crystallographic configuration, the differential docking finds this orientation in 35% less time and eliminates 90% of the undesirable orientations.

DIFFDOCK mode	required sets	# of matches	# of positive scoring orientations	time (sec)	high score	rms of high score
full	none	120833	2587	118	193	0.34
differential	diff	55754	28	77	193	0.34
differential	ser	38884	389	65	193	0.34
differential	diff, ser	19325	6	52	193	0.34

Table 10-3. Docking LPF into chymotrypsin. A full run is compared to differential runs with flagged spheres representing the differential

region alone (diff), the region representing proximity to the hydroxyl oxygen of serine 195 (ser), or both (diff, ser). The full cluster was composed of 83 spheres. Of these, 13 were selected as characterizing the specificity pocket, and 4 lie within 3.5 Å of the OG of serine 195. Input parameters are given in the methods. The number of matches passing the internal distance requirement (see methods) and that use the required spheres are listed. Also reported are the total cpu time for each run in seconds and the root mean square deviation (rms) of all the inhibitor atoms from their crystallographic coordinates.

Additional information about the mode of binding derived from experimental data or other evidence can be used to focus the search further. Trifluoromethyl ketones are known to bind covalently to the hydroxyl oxygen (OG) of serine 195 [27, 37]; therefore, the inhibitor must be bound in the vicinity of that atom. Four spheres in the cluster representing the active site of chymotrypsin lie within 3.5 Å of the OG (Figure 10-2a). Table 10-3 shows the results of DIFFDOCK runs with these four spheres flagged. While docking to the four spheres eliminates 85% of the orientations, requiring each match to use both one of the spheres near the serine and one of the spheres representing the specificity pocket focuses the search to only 6 positive scoring orientations. All 6 have rms within 2 Å of the crystallographic configuration (not shown). In addition to eliminating 99.8% of the undesirable orientations, this run takes only 44% of the time as the full search.

Discussion

DIFFDOCK is a modification of the DOCK program that allows the user to guide the search of orientation space to site regions of particular interest. These regions are user-selected and may represent features discriminating homologous receptors, residues previously determined to participate in binding substrates or inhibitors, or other areas to which a ligand is to be targeted. Narrowing the search significantly decreases run time and eliminates undesired orientations. The increased emphasis on orienting ligands in the most interesting regions of the site is particularly important in database searches since only a single configuration is retained for each small molecule [19].

In the examples presented, the flagged spheres were selected by comparing the crystal structures of each pair of enzymes. Currently this is done graphically by the user although a method is under development to automate identification of spheres representing differential site volume. However, DIFFDOCK does not require three-dimensional structures of both receptors. Site features to be emphasized may be selected based on non-crystallographic information such as sequence conservation or knowledge of functional roles of various residues. By docking to flagged spheres representing shared regions instead of differential regions, DIFFDOCK can be used to identify inhibitors active against multiple receptors. This strategy may also be applied in predicting the binding modes of substrates or cofactors common to two or more enzymes.

While DIFFDOCK was successful in reproducing the crystallographic configurations of the test cases presented, there are several caveats to its general use in predicting binding modes. The conformations used for both ligands and proteins were determined from inhibitor-receptor complexes. Results from docking unbound conformations must be interpreted cautiously as DIFFDOCK treats both molecules as rigid and cannot predict conformational changes induced by binding. Also, while DIFFDOCK can suggest that a ligand will bind to a targeted site it does not consider whether the ligand would preferentially bind to that site over alternate subsites, a rearranged subsite, or pockets on other receptors.

DIFFDOCK is able to identify known selective inhibitors but may generate false negatives. One way in which compounds may be missed is simply due to the fact that the search is nonexhaustive; the heuristics used to prune orientation space [20] potentially eliminate some acceptable orientations in the selected region. Secondly, ligands can dock to the selected region without using a flagged sphere explicitly in the match. Since screening is done at match time before transformation of the ligand into the site, these orientations will be overlooked.

The current implementation of DIFFDOCK in version 2 of DOCK is limited to differentiating two sites spatially since the scoring routine only considers steric interactions [20]. The forthcoming DOCK3 [38] contains a new scoring scheme which includes electrostatic interaction terms from the AMBER potential function [39]. Incorporation of the DIFFDOCK routines into DOCK3 will permit comparison of a wider variety of sites and allow consideration of

ligand atom types. A slight modification to the code will allow selected ligand atoms to be paired with spheres representing particular site features. For example, in the α -lytic protease test case the search could be narrowed further by requiring that only the boron atom of the inhibitor approach the hydroxyl oxygen to which it is covalently bound. Additionally, other points representing site characteristics such as GRID [13] points will be able to direct the docking.

One important application for DIFFDOCK is the design of species specific inhibitors. Drugs need to selectively inhibit the target enzyme of the invading organism without inhibiting the human analog. DIFFDOCK can be used to identify ligands that could potentially bind to regions that discriminate the active sites of the different receptors. Enzymes differing by as little as a single amino acid may be distinguished, as shown by the α -lytic protease example, as can receptors with much lower sequence identity.

While specificity against a known homologous receptor is a crucial property of a good drug, it is not the only qualification. Drugs must be inactive against a large number of proteins, most of which are of unknown structure, and possess features dictating appropriate biological behavior including absorption, targeting, and bioavailability. Although this method does not address these characteristics, incorporation of selectivity against known targets into an inhibitor design strategy is an important first step in creating new safe and effective drugs.

References

1. Bourne, H. R. and Roberts, J. M., "Drug Receptors and Pharmacodynamics." In B. G. Katzung (Ed.), Basic and Clinical Pharmacology. (Appleton & Lange, Norwalk, CT, 1989) pp. 10-28.
2. Korolkovas, A. and Burckhalter, J. H., "Development of Drugs." In Essentials of Medicinal Chemistry. (John Wiley & Sons, New York, 1976) pp. 12-43.
3. Hol, W. G. J., "Protein crystallography and computer graphics - toward rational drug design." *Angew Chem* **25** 767-852 (1986).
4. Martin, Y. C., "Computer-assisted rational drug design." *Meth Enz* **203** 587-613 (1991).
5. Ripka, W. C., Sipio, W. J. and Blaney, J. M., "Molecular modeling and drug design: strategies in the design and synthesis of phospholipase A2 inhibitors." *Lect Heterocyc Chem* **9** S95-S104 (1987).
6. Kuyper, L. F., Roth, B., Baccanari, D. P., Ferone, R., Beddell, C. R., Champness, J. N., Stammers, D. K., Dann, J. G., et al., "Receptor-based design of dihydrofolate reductase inhibitors: comparison of crystallographically determined enzyme binding with enzyme affinity in a series of carboxy-substituted trimethoprim analogs." *J Med Chem* **25** 1120-1122 (1982).
7. DesJarlais, R. L., Seibel, G. L., Kuntz, I. D., Furth, P. S., Alvarez, J. C., Ortiz de Montellano, P. R., DeCamp, D. L., Babe', L. M., et al., "Structure-based design of nonpeptide inhibitors specific for

- the human immunodeficiency virus 1 protease." *PNAS* 87 6644-6648 (1990).
8. Beddell, C. R., Goodford, P. J., Norrington, F. E., Wilkinson, S. and Wootton, R., "Compounds designed to fit a site of known structure in human hemoglobin." *Br J Pharmac* 57 201-209 (1976).
 9. Gilman, A. G., Rall, T. W., Nies, A. S. and Taylor, P. (Eds.), Goodman and Gilman's The Experimental Basis of Therapeutics. (Pergamon Press, New York, 1990).
 10. Krieger, M., Kay, L. M. and Stroud, R. M., "Structure and specific binding of trypsin: comparison of inhibited derivatives and a model for substrate binding." *J Mol Bio* 83 209-230 (1974).
 11. Steitz, T. A., Henderson, R. and Blow, D. M., "Structure of crystalline alpha-chymotrypsin. III. Crystallographic studies of substrates and inhibitors bound to the active site of alpha-chymotrypsin." *J Mol Biol* 46 337-348 (1969).
 12. Shotton, D. M. and Watson, H. C., "Three-dimensional structure of tosyl-elastase." *Nature* 225 811-816 (1970).
 13. Goodford, P. J., "A computational procedure for determining energetically favorable binding sites on biologically important macromolecules." *J Med Chem* 28 849-857 (1985).
 14. Lawrence, M. C. and Davis, P. C., "CLIX: A search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure." *Proteins* 12 31-41 (1992).
 15. Miranker, A. and Karplus, M., "Functionality maps of binding sites: a multiple copy simultaneous search method." *Proteins* 11 29-34 (1991).

16. Van Drie, J. H., Weininger, D. and Martin, Y. C., "ALADDIN: An integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructure searching of three-dimensional molecular structures." *J Comp-Aid Mol Des* 3 225-251 (1989).
17. Moon, J. B. and Howe, W. J., "Computer design of bioactive molecules: a method receptor-based *de novo* ligand design." *Proteins* 11 314-328 (1991).
18. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. and Ferrin, T. E., "A geometric approach to macromolecule-ligand interactions." *J Mol Biol* 161 269-288 (1982).
19. DesJarlais, R. L., Sheridan, R. P., Seibel, G. L., Dixon, J. S., Kuntz, I. D. and Venkataraghavan, R., "Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure." *J Med Chem* 31 722-729 (1988).
20. Shoichet, B. K., Bodian, D. L. and Kuntz, I. D., "Molecular docking using shape descriptors." *J Comp Chem* 13 380-397 (1992).
21. Richards, F. M., "Areas, volumes, packing, and protein structure." *Ann Rev Biophys Bioeng* 6 151-176 (1977).
22. Ferro, D. R. and Hermans, J., "A different best rigid-body molecular fit routine." *Acta Cryst* A33 345-347 (1977).
23. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. and Weng, J., In F. H. Allen, G. Bergerhoff and R. Seivers (Eds.), Crystallographic Databases - Information Content. Software Systems. Scientific Applications. (Data Commission of the

- International Union of Crystallography, Cambridge, 1987) pp. 107-132.
24. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., et al., "The Protein Data Bank: A computer-based archival file for macromolecular structures." *J Mol Biol* 112 535-542 (1977).
 25. Bone, R., Frank, D., Kettner, C. A. and Agard, D. A., "Structural analysis of specificity: α -lytic protease complexes with analogues of reaction intermediates." *Biochem* 28 7600-7609 (1989).
 26. Bone, R., Silen, J. L. and Agard, D. A., "Structural plasticity broadens the specificity of an engineered protease." *Nature* 339 191-195 (1989).
 27. Brady, K., Wei, A., Ringe, D. and Abeles, R. H., "Structure of chymotrypsin-trifluoromethyl ketone inhibitor complexes: comparison of slowly and rapidly equilibrating inhibitors." *Biochem* 29 7600-7607 (1990).
 28. Meyer, E., Cole, G., Radhakrishnan, R. and Epp, O., "Structure of native porcine pancreatic elastase at 1.65 Å resolution." *Acta Cryst B* 44 26-38 (1988).
 29. Connolly, M. L., "Solvent-accessible surfaces of proteins and nucleic acids." *Science* 221 709-713 (1983).
 30. Schechter, I. and Berger, A., "On the size of the active site in proteases. I. Papain." *Biochem Biophys Res Comm* 27 157-162 (1967).
 31. Ferrin, T. E., Huang, C. C., Jarvis, L. E. and Langridge, R., "The Midas display system." *J Mol Graph* 6 13-27 (1988).

32. Bauer, C.-A., Brayer, G. D., Sielecki, A. R. and James, M. N. G., "Active site of alpha-lytic protease. Enzyme-substrate interactions." *Eur J Biochem* 120 289-294 (1981).
33. Kaplan, H. and Whitaker, D. R., "Kinetic properties of the alpha-lytic protease of *Sorangium* sp., a bacterial homologue of the pancreatopeptidases." *Can J Biochem* 47 305-316 (1969).
34. Kettner, C. A., Bone, R., Agard, D. A. and Bachovchin, W. W., "Kinetic properties of the binding of α -lytic protease to peptide boronic acids." *Biochem* 27 7682-7688 (1988).
35. Shotton, D. M. and Hartley, B. S., "Amino acid sequence of porcine pancreatic elastase and its homology with other serine proteinases." *Nature* 225 802-806 (1970).
36. Imperiali, B. and Abeles, R. H., "Inhibition of serine proteases by peptidyl fluoromethyl ketones." *Biochem* 25 3760-3767 (1986).
37. Liang, T.-C. and Abeles, R. H., "Complex of alpha-chymotrypsin and N-acetyl-L-leucyl-L-phenylalanyl trifluoromethyl ketone: structural studies with NMR spectroscopy." *Biochem* 26 7603-7608 (1987).
38. Meng, E. C., Shoichet, B. K. and Kuntz, I. D., "Automated docking with grid-based energy evaluation." *J Comp Chem* 13 505-524 (1992).
39. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr and Weiner, P., "A new force field for molecular mechanical simulation of nucleic acids and proteins." *JACS* 106 765-784 (1984).

Appendix A

Chemical Synthesis and Characterization

The following information is provided by the chemists at SERES Labs who did the synthesis and characterization of both designed and commercially available compounds. Synthesis and characterization of the adamantyl derivatives adm1-3, 6-8 were done by Patricia Caldera. Erin Bradley took NMR spectra of compounds adm3 and njo8.

Synthesis of Round I Compounds

njo6: 3',5'-O,O-biscarboxymethyl-N⁴-benzoyl-2'-deoxycytidine. N⁴-benzoyl-2'-deoxycytidine (200 mg, 0.604 mmol) was suspended in 10 ml of tetrahydrofuran under dry nitrogen atmosphere and 297 mg (2.66 mmol) of potassium tert-butoxide was added. This mixture was stirred for one hr at room temperature, then heated at reflux for an additional hr. Solid chloroacetic acid (114 mg, 1.21 mmol) was then added and the resulting suspension was refluxed for 1 hr. After cooling to 0°, the mixture was acidified with hydrogen chloride gas and the resulting solid was collected by filtration and washed with acetone to give an off-white solid (290 mg). This solid begins to char at 120° but does not melt below 300°; IR (KBr, cm⁻¹) 3500-2880, 1740, 1685, 1605. Anal. calculated for C₂₀H₂₁N₃O₉: C, 53.69; H 4.70; N, 9.40. Found: C, 20.17, H 1.99, N, 4.34. This material was used without further purification.

njo7: 3',5'-O,O-biscarboxymethyl-2'-deoxycytidine. The N-benzoyl precursor njo6 (180 mg, 0.40 mmol) was dissolved in 3 ml of

concentrated ammonium hydroxide and refluxed for 30 min; an additional 3 ml of ammonium hydroxide was then added and refluxing was continued for another 30 min. The aqueous solution was evaporated under vacuum, and the resulting solid was dissolved in 2 ml of water. The pH of this solution was carefully adjusted to 5.45 with 12 N hydrochloric acid. The solution was then evaporated under vacuum and the resulting solid (ca. 100 mg) was washed with acetone. This solid was then dissolved in ca 0.5 ml of distilled water and purified by HPLC at 50° on a Spherisorb ODS-2, 10 micron preparative column (Column Engineering Inc., Ontario, CA) at a flow rate (water) of 6 ml/min. Peaks were detected by absorption at 200 nm. The peak eluting at 14 min, major fraction, was collected and shown to be homogeneous on a similar analytical column (DeltaBond 300, 5 micron Octyl, Keystone Scientific, State College, PA). IR (KBr, cm^{-1}) 3500-2700, 1660, 1610, 1550. The non-integrated, 500 Hz nmr spectrum is consistent with the proposed structure. However, further studies indicated that the final product was predominantly 2'-deoxycytidine.

njo8: N^4 -benzoyl-3',5'-O,O-biscarbamyl-2'-deoxycytidine. To a solution of 3'-(di(4-methoxyphenyl)-phenylmethyl) N^4 -benzoyl-2'-deoxycytidine and quinoline in 50 ml of dry tetrahydrofuran (THF) at 0°C was added trichloromethylchloroformate. The reaction mixture was allowed to warm to 25 °C and stirred for 24 hours. The THF soluble material was cannulated to a clean dry flask. Gaseous ammonia was bubbled through the solution for 5 mins. The product was isolated by filtration and purified by RP-HPLC (column temperature = 45 °C; flow rate 3.5 ml/min; gradient = 1 over 15 minutes, solvent 20-80% methanol-water). A white solid was collected at R_t = 14.5 mins. Mass spectrum (LSIMS) m/e 418 $[\text{MH}]^+$, 440 $[\text{MNa}]^+$, 837 $[\text{M}(\text{dimer})]^+$, 859 $[\text{M}(\text{dimer})\text{Na}]^+$.

njo9: 2'-deoxycytidine-3',5'-dicarbamate. To a stirred solution of 700 mg (1.1 mmol) of 5'-O-(4,4'-dimethoxytrityl)- N^4 -benzoyl-2'-deoxycytidine and 0.25 ml of trichloromethyl chloroformate in 50 ml of dry THF at 25 °C was added 0.25 ml of quinoline. The resulting mixture was allowed to stir for 2 h before transferring the soluble portion into a clean vessel and cooling to 0 °C. Excess ammonia gas was then bubbled into the stirred reaction solution. The suspension that resulted was stirred an additional 60 mins before evaporating to dryness *in vacuo*. The residue was dissolved in water, neutralized (pH = 7) with dilute hydrochloric acid, and dried again. The resulting solid was added to 10 ml of 50% aqueous acetic acid and the

suspension was stirred at 25 °C for 60 min. The mixture become homogeneous and was evaporated *in vacuo* to afford a yellow paste. The paste was dissolved in methylene chloride and extracted with 50% aqueous methanol. The methanolic solution was concentrated *in vacuo* to afford a pale brown solid. Purification by RP-HPLC (column temperature = 45 °C; flow rate 3.0 ml/min; gradient = 1 over 15 minutes, 20-80% methanol-water) afforded a white solid which was then dissolved in 50% aqueous ammonium hydroxide and heated at reflux for 1 h. The solvent was removed *in vacuo* to afford a pale brown solid. Purification by RP-HPLC (column temperature = 40 °C; flow rate 3.0 ml/min; gradient = 1 over 15 minutes, 26-100% methanol-water) afforded 640 mg of a pale yellow solid. Final purification by RP-HPLC (column temperature = 25 °C; flow rate 2.2 ml/min; gradient = 2 over 8 minutes, 4-40% methanol-water) afforded 10 mg of white solid. Mass spectrum (LSIMS) m/e 313 [MH]⁺, 336 [MNa]⁺.

njo12: 2'-deoxycytidine-3',5'-monocarbamate. This compound was isolated during the final purification of njo9. (LSIMS) m/e 271 [MH]⁺.

njo14: Isolated as an intermediated during the synthesis of njo9.

njo26: 5'-O-carboxymethyl-2'-deoxycytidine. 400 mg (100 mmol) of sodium hydroxide was added to dry dimethylsulfoxide at room temperature and stirred for 15 minutes. Slowly, 441 mg (4.8 mmol) of chloroacetic acid was added to this mixture and stirred. After 15 minutes 530 mg (1.6 mmol) of N⁴-benzoyl-2'-deoxycytidine was added and the resulting mixture was stirred for 70 hours. The reaction was added and the resulting mixture was stirred for 70 hours. The reaction solution was concentrated to dryness *in vacuo*. The resulting residue was taken up in 20 ml of concentrated ammonium hydroxide and heated for 90 minutes. Evaporation of the solvent *in vacuo* afforded a pale brown solid. Purification by RP-HPLC (column temperature = 25 °C; flow rate 2.2 ml/min; gradient = 2 over 15 minutes, 4-80% methanol water) afforded a 70 mg hygroscopic white solid (R_t = 7.9 mins). Mass spectrum (LSIMS) m/e 344 [MH]⁺, 366 [MNa]⁺, 388 [MNa₂]⁺, 410 [MNa₃]⁺.

njo27: 2'-deoxycytidine-5'-carbamate. To a stirred solution of 1.0 g of N⁴-benzoyl-2'-deoxycytidine and 1.0 ml of trichloromethyl chloroformate in 40 ml of dry THF at 25 °C was added 1.0 ml of quinoline. The resulting mixture was allowed to stir for 3.5 h. The

THF-soluble material was cannulated into a clean vessel and cooled to 0 °C. Into the stirred THF solution was bubbled excess ammonia gas. The resulting suspension was stirred an additional 30 min and evaporated *in vacuo* to afford a pale brown solid, which after purification by RP-HPLC (column temperature = 40 °C; flow rate 3.0 ml/min; gradient = 1 over 15 minutes, 26-100% methanol water) afforded a 640 mg of a pale yellow solid. Further purification by RP-HPLC (column temperature = 25 °C; flow rate 2.2 ml/min; gradient = 2 over 8 minutes, 4-40% methanol water) afforded a white solid. LSIMS m/e 271 [MH]⁺, 293 [MNa]⁺.

ncco5: 3-carboxymethyl-1-amantanol. A solution of 6.1 g (21.2 mmol) of methyl 3-bromo-1-adamantaneacetate in 10 ml of methanol was mixed with 6.27 g of sodium hydroxide in 20 ml of water. The reaction mixture was refluxed for 30 min; methanol was then removed under vacuum. The aqueous solution was washed three times with methylene chloride. The extract was dried over sodium sulfate, and concentrated to produce white crystals. These were recrystallized from benzene to produce 2.33 g (52%) of product acid, mp 116-120°. Anal. Calc'd for C₁₂H₁₈O₃: C, 68.57%; H 8.57%. Found: C, 68.45%, H, 8.71%.

ncco6: 3-(2-hydroxyethyl)-1-amantanol. A solution of 190 mg (1.06 mmol) 3-carboxymethyl-1-amantanol in 5 ml of dry tetrahydrofuran under a nitrogen atmosphere was cooled to 0°, and 1.42 ml of a 1 M solution of borane in tetrahydrofuran was added. The solution was allowed to reach room temperature with stirring. After one hr, an excess of 1:1 tetrahydrofuran-water was added dropwise to destroy unreacted borane. The mixture was then treated with saturated potassium carbonate, and extracted twice with ethyl acetate. The organic phase was dried over sodium sulfate and evaporated to yield 130 mg (70%) of a white solid mp 57-67°. Recrystallization from benzene produced material melting at 92-95°. Anal. Calc'd for C₁₂H₂₀O₂: C, 73.47%, H 10.20%. Found: C, 72.69%, H, 10.23%.

3-(2-tosyloxyethyl)-1-amantanol. 3-(2-hydroxyethyl)-1-amantanol (100 mg, 0.510 mmol) was dissolved in 1 ml of pyridine and cooled in an ice bath; 98 mg (0.512 mmol) of p-toluenesulfonyl chloride, which had been previously recrystallized from hexane, was added and the mixture was refrigerated overnight. The reaction mixture was washed with 2 N hydrochloric acid until the washings were acidic. The organic layer was then washed with water until neutral and dried over sodium sulfate. Evaporation under vacuum produced

a colorless oil with was recrystallized from ether-hexane to give white crystals (88 mg, 49%) mp 67-68°. Anal. Calc'd for $C_{19}H_{26}O_4S$: C, 65.14%, H, 7.43%, S, 9.14%. Found: C, 65.34%, H, 7.67%, S, 8.89%.

3-(2-cyanoethyl)-1-amantanol. 3-(2-tosyloxyethyl)-1-amantanol (500 mg, 1.4 mmol) was dissolved in 5 ml of dimethyl sulfoxide and 200 mg (4.0 mmol) of sodium cyanide was added. The stirred reaction mixture was heated until the sodium cyanide was dissolved, and then allowed to proceed at room temperature for 21 hr. The reaction mixture was extracted three times with methylene chloride. The extract was washed three times with water and then dried over sodium sulfate. Evaporation under vacuum produced a pale yellow solid (260 mg, 90.6%) which was recrystallized from benzene-hexane to give white crystals, mp 82-84°. Anal. Calc'd for $C_{13}H_{19}ON$: C, 76.10%, H, 9.27%, N, 6.83%. Found: C, 76.06%, H, 9.41%, N, 6.82%.

ncco7: 3-(2-cyanoethyl)-N-acetylamantadine. A procedure similar to that published was used to introduce the acetamido group [1]. A solution of 780 mg (3.80 mmol) of 3-(2-cyanoethyl)-1-amantanol in 31 ml of acetonitrile under nitrogen was cooled to 0°. Sulfuric acid, 12 ml of an 18 M solution, was added dropwise after which the reaction mixture was allowed to come to room temperature. After 2.5 hr, an excess of solid sodium hydroxide was slowly added. The mixture was then extracted three times with methylene chloride and then dried over sodium sulfate. The product was purified by flash chromatography on silica gel (1:1 acetone-methylene chloride eluent). The resulting white solid (633 mg, 68%) was recrystallized from methylene chloride-hexane to give material melting at 110-111°. Anal. Calc'd for $C_{15}H_{22}N_2O$: C, 73.17%, H, 8.94%, N, 11.38%. Found: C, 72.58%, H, 9.15%, N, 11.08%.

ncco8 (adm8): 3-(2-carboxyethyl)-1-amantadine hydrochloride. A suspension of 3-(2-cyanoethyl)-N-acetylamantadine (402 mg) in 25 ml of 6N hydrochloric acid was refluxed under nitrogen for 45 hr. The cooled solution was filtered to remove a white precipitate. The aqueous solution was evaporated under vacuum to produce a pale yellow solid (195 mg, 46%) which was washed with acetone. Anal. Calc'd for $C_{13}H_{22}NO_2Cl$: C, 60.12%, H, 8.48%, N, 5.39%, Cl, 13.68%. Found: C, 14.88%, H, 7.84%, N, 20.09%. HPLC of this product on DeltaBond 300, 5 micron Octyl (Keystone Scientific, State College, PA) showed two peaks (2.38 min and 7.08 min) when eluted with 1:10 acetonitrile water. Peaks were detected by refractive index monitoring. The non-integrated 500 Hz NMR spectrum was

consistent with the assigned structure. IR (KBr, cm^{-1}) 3400-2100, 1650. Further analysis revealed the product to be predominantly ammonium chloride.

adm2 (adm1) 3-amino-1-adamantaneethanol. This compound was made in two steps by first converting methyl 3-bromo-1-adamantane acetate into methyl 3-azido-1-adamantaneacetate and then reducing the azido group and methyl ester to yield the product. A solution of methyl 3-bromo-1-adamantane acetate (1.34 g, 4.6 mmol) in 10 ml of dichloromethane was cooled to 0° in an ice bath. Trimethylsilylazide (0.74 ml, 5.5 mmol) was added followed by 4.8 ml (4.8 mmol) of 1 M titanium chloride IV in dichloromethane. The cooling bath was removed and the reaction mixture was stirred for 16 h before being poured into a cold solution of 2% sodium carbonate. The resulting suspension was extracted with dichloromethane. The organic extracts were combined, washed with water and brine, and dried over sodium sulfate. Evaporation of the solvent yielded a syrup-like residue (1.04 g). The infrared spectrum showed the azido group absorption at 2087 cm^{-1} . A solution of the azide (830 mg, 3.3 mmol) in 10 ml of diethyl ether was added dropwise (to maintain a gentle reflux) to a suspension of lithium aluminum hydride (170 mg) in 20 ml diethyl ether. The reaction mixture was heated at reflux for 2 h, cooled, and quenched by slow addition of a saturated solution of ammonium sulfate until a precipitate formed. Filtration left a clear solution, which on alumina thin-layer chromatography (chloroform-methanol 9:1) showed two spots, a less polar one ($R_f = 0.68$) corresponding to sideproduct 3-chloro-1-adamantaneethanol and a more polar one ($R_f = 0.21$) corresponding to the desired product. The two components were separated by bubbling hydrogen chloride gas into the ethereal filtrate and collecting the partially crystalline precipitate. The precipitate was recrystallized from methanol-diethyl ether to produce the hydrochloride salt as white crystals (183 mg): mp $236-237^\circ\text{C}$; proton NMR (D_2O) 3.67 (t, 2H, $-\text{CH}_2\text{OH}$), 2.4-1.4 (m, 18H); MS (free base) m/e 195 (M^+ 38%), 138 ($\text{M}-\text{C}_4\text{H}_9$, 100%); elemental analysis, calc'd: C 62.18%, H 9.56%, N 6.04%. Found: C 61.96%, H 9.47%, N 5.97%.

adm3 3-amino-1-adamantaneacetic acid. A suspension of methyl-3-amino-1-adamantaneacetate (470 mg, 2.10 mmol) in 6 ml of 20% HCl was refluxed for 20 hr. The cooled suspension was washed with ether, which caused the product to crystallize. The crystals were filtered and washed with ether containing a few drops of methanol

and recrystallized from methanol/ether. m.p. 230-232 °C (250 mg, 48% yield); HNMR (D₂O): δ 2.4-1.4 (m, 18H); MS: m/e 210 (M⁺).

adm6 3-amino-1-adamantane(3-propanol). Prepared as 3-amino-1-adamantaneethanol from the azide described above.(284 mg, 1.08 mmol) yielded 109 mg (52%) of white crystals. Conversion to the hydrochloride salt yielded crystalline material which was recrystallized from methanol/ether in 70% yield. m.p. 232-233 °C. HNMR (D₂O): δ 3.6 (t, 2H, CH₂-OH), 2.4-1.1 (m, 20H); MS: m/e 209 (M⁺, 28%), 152 (M⁺ - C₄H₉, 100%); microanalysis Calc. C 63.53%, H 9.84%, N 5.70%; found C 62.8%, H 9.57%, N 5.59%).

adm7 methyl-3-amino-1-adamantaneacetate. To the crude azide (1.44 g, 5.7 mmol) in 10 ml of methanol was added 240 mg of Mg⁰ turnings. To initiate the reaction the mixture was warmed up gently until bubbling was observed. The progress of the reaction was monitored by TLC. After all the magnesium had dissolved, the suspension was filtered and the precipitate washed with methanol. The methanol was evaporated and the residue was suspended in cold water. The solution was adjusted to pH 9-10 with 5% sodium hydroxide and extracted with ether. The ether layer was then extracted with 5% HCl. The aqueous solution was then basified again and extracted with ether. The organic layer was washed with water and brine, dried over sodium sulfate. Evaporation of solvent yielded a yellow syrup-like residue (504 mg, 40% yield). IR: 3400 cm⁻¹ (broad, NH₂); 1735 cm⁻¹ (sharp, carbonyl); HNMR (CDCl₃): δ 3.62 (s, 3H, COCH₃), 2.18 (s, 2H, CH₂CO), 1.8 - 1.3 (m, 16H); MS: m/e 223 (M⁺, 18%), 166 (M⁺ - C₄H₉, 100%). A portion of the product was converted to the hydrochloride salt and recovered as white crystals (m.p. 195 °C).

Purification and Synthesis of Derivatives of Compound 83

83K 4A.5.8.8A-tetrahydro-5.8-methano-1.4-naphthoquinone. The commercial compound was recrystallized twice from hexane. The structure was confirmed by H-NMR and IR. In the IR there were no OH bands but strong carbonyl bands centered at 1600 cm⁻¹. Elem Anal Calc'd: C 75.84%, H 5.79%. Found: C 75.75%, H 5.80%.

83A 5.8-dihydro-5.8-methano-1.4-naphthalenediol. Prepared from 83K by method of Porter, et. al. [2]. The structure was confirmed by H-NMR and IR (no carbonyl bands but strong OH band). Elem anal Calc'd: C 75.84%, H 5.79%. Found: C 75.62%, H 5.66%.

compound	purification	elem anal	mp (°C)	lit mp (°C)	NMR
90	recrys Et ₂ O	passed	124-125	124-125	ok
99	sublim	passed	223-225	223-225	ok
111	as is	passed	178-180	180	ok
114	sublim	passed	218	216-218	ok
116	sublim	passed	216.5- 217.5 sub	217 sub	ok
117	as is	passed	126.5- 128	127-129	ok
118	as is	passed	105.5- 106.5	107	ok
119	recrys CHCl ₃ - EtOH	passed	160-163	161-163	ok
120	recryst. EtOH - HOAc	passed	193-195	191-192	ok
121	recryst. HCl(aq)-SnCl ₂	passed as 1/4 H ₂ O	--		ok
125	recryst. HCl(aq)-SnCl ₂	failed twice	236	--	ok
126	recryst. MeNO ₂	passed	260-262 dec	259-261 dec	ok

Table A-1. Summary of purification and characterization of several derivatives of compound 83. Literature and experimental melting points are compared. The method of purification, and summaries of the elemental analysis and NMR experiments are given. Details of the elemental analysis are in Table A-2.

Structures of compounds 101, 102, 103, 104, 108 were confirmed by H-NMR. TLC analysis showed commercial samples of 91 and 97 are complex mixtures.

compound	calculated	found
90	C 75.94%, H 3.82%	C 75.84%, H 3.94%
99	C 69.54%, H 7.30%	C 69.78%, H 7.32%
111	C 74.97%, H 6.86%	C 74.77%, H 7.03%
114	C 80.75%, H 3.87%	C 80.99%, H 3.97%
116	C 69.54%, H 7.30%	C 69.78%, H 7.32%
117	C 72.25%, H 8.49%	C 72.46%, H 8.55%
118	C 76.73%, H 4.68%	C 76.60%, H 4.70%
119	C 68.96%, H 3.47%	C 69.19%, H 3.60%
120	C 68.96%, H 3.47%	C 68.96%, H 3.59%
121 ($\cdot 1/4$ H ₂ O)	C 60.00%, H 5.30%, N 7.00%, Cl 17.71%	C 60.39%, H 5.27%, N 6.67%, Cl 17.52%
		C 60.23%, H 5.28%, N 6.59%
125	--	--
126	C 74.98%, H 5.03%	C 74.67%, H 5.15%

Table A-2. Summary of elemental analysis results. The experiment was repeated twice with compound 121. Compound 125 failed this test.

References

1. Lenoir, D., Glaser, R., Mison, P. and von Rague Schleyer, P., "Synthesis of 1,2- and 2,4-disubstituted adamantanes. The protoadamantane route." *J Org Chem* 36 1821 (1971).
2. Porter, R. F., Rees, W. W., Frauenglass, E., Wilgus, H. S., Nawn, G. H., Chiesa, P. P. and Gates, J. W., "The chemistry of thioether-substituted hydroquinones and quinones. I. The 1,4 addition of a heterocyclic mercaptan to quinones." *J Org Chem* 29 588-594 (1964).

Appendix B

Hemagglutinin Sequences

Hemagglutinin sequences were aligned following Kawaoka, et al. [1]. The one letter amino acid code was used. '.' means that residue was not sequenced, and '-' represents a gap introduced for optimal alignment. Inconsistencies were resolved by selecting published sequences over GCG databank entries [2]. Sequences included, their abbreviations and their sources are:

H3 sequences:

H3	X31	X31 [3]
H3	Pc173	A/PC/1/73 [4]
H3	Eng4272	A/England/42/72 [4]
H3	Hk10771	A/Hong Kong/107/71 [4]
H3	Qu770	A/Queensland/7/70 [4]
H3	Eng69	A/England/69 [4]
H3	Nt68	A/NT/60/68 [4]
H3	X31w68	X31 mutant [4]
H3	Vic375	A/Victoria/3/75 [5]
H3	Bk179	A/Bangkok/1/79 [6]
H3	Bk279	A/Bangkok/2/79 [7]
H3	Lenin38580	A/Leningrad/385/80 [7]
H3	Eng32177	A/England/321/77 [8]
H3	Mem171	A/Memphis/1/71 [9]
H3	Mem10272	A/Memphis/102/72 [9]
H3	Aichi268	A/Aichi/2/68 (GCG hmivh)
H3	Nt606829c	A/NT/60/68/29C [6]
H3	Tex77	A/Tex/77 [10]
H3	Mem10772	A/Mem/107/72 [9]
H3	Duokr63	A/Duck/Ukraine/63 (GCG hmivdu)

H3 EQ urg63	A/Equine/Uruguay/1/63 [11]
H3 EQ mia63	A/Equine/Miami/1/63 [11]
H3 EQ tky71	A/Equine/Tokyo/71 [11]
H3 EQ alg72	A/Equine/Algiers/72 [11]
H3 EQ nm 76	A/Equine/New Market/76 [11]
H3 EQ fon79	A/Equine/Fontainebleau/79 [11]
H3 EQ rom80	A/Equine/Romania/80 [11]
H3 EQ san85	A/Equine/Santiago/1/85 [11]
H3 EQ tn 86	A/Equine/Tennessee/5/86 [11]
H3 EQ ky 86	A/Equine/Kentucky/2/86 [11]
H3 EQ ky 87	A/Equine/Kentucky/1/87 [11]
H3 DU hk577	A/Duck/Hokkaido/5/77 [12]
H3 DU hk880	A/Duck/Hokkaido/8/80 [12]
H3 DU hk3380	A/Duck/Hokkaido/33/80 [12]
H3 DU hk782	A/Duck/Hokkaido/7/82 [12]
H3 DU hk2182	A/Duck/Hokkaido/21/82 [12]
H3 DU hk985	A/Duck/Hokkaido/9/82 [12]
H3 DU hk1085	A/Duck/Hokkaido/10/82 [12]
H3 SW hk12682	A/Swine/HongKong/126/82 [13]
H3 SW hk8178	A/Swine/HongKong/81/78 [13]

H4 sequences:

H4 DU Czk56	A/Duck/Czeckoslovakia/56 [14]
H4 GT Aus79	A/Grey Teal/Australia/2/79 [14]
H4 BU Hok77	A/Budgerigar/Hokkaido/1/77 [14]
H4 DU NZ76	A/Duck/New Zealand/31/76 [14]
H4 DU Alb76	A/Duck/Alberta/28/76 [14]
H4 CH Ala75	A/Chicken/Alabama/1/75 [14]
H4 RT NJ85	A/Ruddy Turnstone/NJ/47/85 [14]
H4 TU Min80	A/Turkey/Minnesota/833/80 [14]
H4 SE Mas82	A/Seal/Massachusetts/133/82 [14]

H1 sequences:

H1 Khab77	A/Khabarovsk/77 [15]
H1 Kiev5979	A/Kiev/59/79 [15]
H1 Wsn33	A/WSN/33 (GCG hmiv5)
H1 Pr832	A/Pr/8/32 [9]
H1 Mem1078	A/Mem/10/78 ("USSR") [9]
H1 Pr834	A/PR/8/34 (Cambridge) [16]
H1 Swnj1176	A/swine/NJ/11/76 (GCG hmivn1)
H1 H1hmiv	A/PR/8/34 (Mount Sinai) (GCG hmiv) [17]

H1 Taiw186	A/Taiwan/1/86 [18]
H1 Chile183	A/Chile/1/83 [18]
H1 HK3283	A/Hong Kong/32/83 [18]
H1 HK282	A/Hong Kong/2/82 [18]

H2 sequences:

H2 Jap30757	A/JAP/307/57 [9]
H2 Ri5-57	A/RI/5-/57 (GCG hmivh2)
H2 Jap30557	A/Japan/305/57 (GCG hmiv2)

H5 sequences:

H5 Chpa83	A/Chicken/Penn/83 [1]
H5 Shaus75	A/shearwater/Australia/75 (GCG hmivh5)
H5 TUIre137883	A/Turkey/Ireland/1378/83 [19]
H5 DUIre11383	A/Duck/Ireland/113/83 [19]

H6, H8, H9 sequences:

H6 Shaus72	A/shearwater/Australia/72 (GCG hmivh6)
H8 Tuont68	A/turkey/Ontario/6118/68 (GCG hmivhb)
H9Tuwisc166	A/turkey/Wisconsin/1/66 (GCG hmivh9)

H10 sequences:

H10Duckman53	A/duck/Manitoba/53 (GCG hmiv10)
H10Minksw84	A/mink/Sweden/84 [20]
H10Chgern49	A/chick/Germany/N/49 [20]

H12 sequence:

H12Dualb6076	A/Duck/Alberta/60/76 (GCG hmiv12)
--------------	-----------------------------------

H7 sequences:

H7 Stvic85	A/Starling/Victoria/5156/85 [21]
H7 Chvic85	A/Chicken/Victoria/1/85 [21]
H7Sealma80	A/Seal/Mass/1/80 [22]
H7 Tuore71	A/Turkey/Oregon/71 (GCG hmivh7)
H7 Fpv	Fowl Plague Virus (GCG hmivf)

H11 sequences:

H11 Duny78
H11Ternaus75
H11Duukr60
H11Dueng56
H11Dumem76

A/Duck/New York/12/78 [23]
A/Tern/Australia/75 [23]
A/Duck/Ukraine/60 [23]
A/Duck/England/56 [23]
A/Duck/Memphis/546/76 (GCG hmiv11)

H3	X31	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	Pc173	QNLPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Eng4272	QDLPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Hk10771	QDLPGNDN-SKATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	Qu770	QDLPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Eng69	QDLPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Nt 68	QDLPGNDNN-TATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	X31w68	QYLPGNPN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Vic375	QDLPGNDNNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Bk179	QNLPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Bk279	QNLPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Lenin38580	QNLPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Eng32177	QNLPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Mem171	QYLPGNPN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Mem10272	QDFPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Aichi268	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	Nt 606829c	QDLPGNDNN-TATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	Tex77	QNLPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Mem10772	QDFPGNDN-STATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKIC
H3	Duukr63	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	EQ urg63	SQNP IGGKN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSTGKIC
H3	EQ mia63	SQNPTGGNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSTGKIC
H3	EQ tky71	SQIPINDNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSTGKIC
H3	EQ alg72	SQIPISDNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSIGKIC
H3	EQ nm 76	SQNPISGNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSIGKIC
H3	EQ fon79	SQNPTSGNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSIGKIC
H3	EQ rom80	SQNPTSGNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSIGKIC
H3	EQ san85	SQNPTSGNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSIGKIC
H3	EQ tn 86	SQNPTSGNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSIGKIC
H3	EQ ky 86	SQNPTSGNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSIGKIC
H3	EQ ky 87	SQNPTSGNN-TATLCLGHHAVANGTLVKTITDDQIEVTNATELVQSTSIGKIC
H3	DU hk577	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	DU hk880	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	DU hk3380	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	DU hk782	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	DU hk2182	QDYSENNN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	DU hk985	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	DU hk1085	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	SW hk12682	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H3	SW hk8178	QDLPGNDN-STATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSSTGKIC
H4	DU Czk56	QNYTGPNVICLGHHAVANGTMVKTLADDQVEVVTAQELVESQNLPELC
H4	GT Aus79	QNYTGPNVICLGHHAVANGTMVKTLADDQVEVVTAQELVESQNLPELC
H4	BU Hok77	QSYTGPNVICLGHHAVANGTMVKTLADDQVEVVTAQELVESQNLPELC
H4	DU NZ76	QNYTGPNVICLGHHAVANGTMVKTLADDQVEVVTAQELVESQNLPELC
H4	DU Alb76	QNYTGPNVICLGHHAVSNGTLVKTITDDQVEVVTAQELVESQHLPELC
H4	CH Ala75	QNYTGPNVICLGHHAVSNGTLVKTITDDQVEVVTAQELVESQHLPELC

H3	X31
H3	Pc173
H3	Eng4272
H3	Hk10771
H3	Qu770
H3	Eng69
H3	Nt68
H3	X31w68
H3	Vic375
H3	Bk179
H3	Bk279
H3	Lenin38580
H3	Eng32177
H3	Mem171
H3	Mem10272
H3	Aichi268
H3	Nt606829c
H3	Tex77

N-NPHRILDGIDCTLIDALLGDPHCDVF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGINCTLIDALLGDPHCDGF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGIDCTLIDALLGDPHCDGF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGIDCTLIDALLGDPHCDVF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGIDCTLIDALLGDPHCDVF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGINCTLIDALLGDPHCDVF-QDETWDLFVERSKA-FSNCYP
N-NPHRILDGIDCTLIDALLGDPHCDVF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGIDCTLIDALLGDPHCDGF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGINCTLIDALLGDPHCDGF-QNEKWDLFVERSKA-FSNCYP
DS-PHRILDGKNCTLIDALLGDPHCDGF-QNEKWDLFVERSKA-FSNCYP
DS-PHRILDGKNCTLIDALLGDPHCDGF-QNEKWDLFVERSKA-FSNCYP
YS-PHRILDGKNCTLIDALLGDPHCDGF-QNEKWDLFVERSKA-FSNCYP
DS-PHRILDGKNCTLIDALLGDPHCDGF-QNEKWDLFVERSKA-FSNCYP
N-NPHRILDGIDCTLIDALLGDPHCDGF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGIDCTLIDALLGDPHCDGF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGIDCTLIDALLGDPHCDVF-QNETWDLFVERSKA-FSNCYP
N-NPHRILDGIDCTLIDALLGDPHCDVF-QNETWDLFVERSKA-FSNCYP
DS-PHRILDGKNCTLIDALLGDPHCDGF-QNEKWDLFVERSKA-FSNCYP

1000
1000
1000
1000
1000

1000

1000

1000

1000

1000

1000

1000

1000

H3	Mem10772	N-NPHRILDGIDCTLIDALLGDPHCDGF-QNETWDLFVERSNA-FSNCYP
H3	Duukr63	N-NPHRILDGRCTLIDALLGDPHCDVF-QNETWDLFVERSNA-FSNCYP
H3	EQ urg63	N-NPYRVLDGRNCTLIDAMLGDPHCDVF-QYGNWDLFIERSSA-FSNCYP
H3	EQ mia63	N-NPYRVLDGRNCTLIDAMLGDPHYDVF-QYENWDLFIERSSA-FSNCYP
H3	EQ tky71	N-NSYRVLDGKNCTLIDAMLGDPHCDVF-QYKNWDLFVERSNA-FSNCYP
H3	EQ alg72	N-NPYRVLDGKNCTLIDAMLGDPHCDVF-QYEKWDLFVERSNA-FSNCYP
H3	EQ nm 76	N-NPYRVLDGRNCTLIDAMLGDPHCDVF-QYENWDLFIERSSA-FSNCYP
H3	EQ fon79	N-NPYRVLDGRNCTLIDAMLGDPHCDVF-QYENWDLFIERSSA-FSNCYP
H3	EQ rom80	N-NPYRVLDGRNCTLIDAMLGDPHCDVF-QYENWDLFIERSSA-FSNCYP
H3	EQ san85	N-NPYRVLDGRNCTLIDAMLGSPHCDVF-QYENWDLFIERSSA-FSNCYP
H3	EQ tn 86	N-NSYRVLDGRNCTLIDAMLGDPHCDVF-QYENWDLFIERSSA-FSNCYP
H3	EQ ky 86	N-NSYRVLDGRNCTLIDAMLGDPHCDVF-QYENWDLFIERSSA-FSNCYP
H3	EQ ky 87	N-NSYRVLDGRNCTLIDAMLGDPHCDVF-QYENWDLFIERSSA-FSNCYP
H3	DU hk577	N-NPHRILDGRDCTLIDALLGDPHCDVF-QDETWDLFVERSNA-FSNCYP
H3	DU hk880	N-NPHRILDGRDCTLIDALLGDPHCDVF-QDETWDLFVERSNA-FSNCYP
H3	DU hk3380	N-NPHRILDGRDCTLIDALLGDPHCDVF-QDETWDLFVERSNA-FSNCYP
H3	DU hk782	N-NPHRILDGRDCTLIDALLGDPHCDVF-QDETWDLFVERSNA-FSNCYP
H3	DU hk2182	N-NPHRILDGRDCTLIDALLGDPHCDVF-QDETWDLFVERSNA-FSNCYP
H3	DU hk985	N-NPHRILDGRDCTLIDALLGDPHCDVF-QDETWDLFVERSNA-FSNCYP
H3	DU hk1085	N-NPHRILDGRDCTLIDALLGDPHCDVF-QDETWDLFVERSNA-FSNCYP
H3	SW hk12682	N-NPHKILDGRDCTLIDALLGDPHCDVF-QDETWDLFVERSNA-FSNCYP
H3	SW hk8178	N-NPHKILDGIDCTLIDALLGDPHCDVF-QDETWDLFVERSNA-FSNCYP
H4	DU Czk56	PS-PLRLVDGQTCIDIINGALGSPGCDHLNGAEDVFI-ERPNA-VDTCTYP
H4	GT Aus79	PS-PLRLVDGQTCIDIINGALGSPGCDHLNGAEDVFI-ERPNA-VDTCTYP
H4	BU Hok77	PS-PLRLVDGQTCIDIINGALGSPGCDHLNGAEDVFI-ERPNA-VDTCTYP
H4	DU N276	PS-PLRLVDGQTCIDIINGALGSPGCDHLNGAEDVFI-ERPNA-VDTCTYP
H4	DU Alb76	PS-PLRLVDGQTCIDIINGALGSPGCDHLNGAEDVFI-ERPNA-VDTCTYP
H4	CH Ala75	PS-PLRLVDGQTCIDIINGALGSPGCDHLNGAEDVFI-ERPNA-VDTCTYP
H4	RT NJ85	PS-PLRLVDGQTCIDIINGALGSPGCDHLNGAEDVFI-ERPNA-VDTCTYP
H4	TU Min80	PS-PLRLVDGQTCIDIINGALGSPGCDHLNGAEDVFI-ERPNA-VDTCTYP
H4	SE Mas82	PS-PLRLVDGQTCIDIINGALGSPGCDHLNGAEDVFI-ERPNA-VDTCTYP
H1	Khab77	RLKGIAPLQLGKCSIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	H1Kiev5979	RLKGIAPLQLGKCSIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	Wsn33	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	Pr832	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	Mem1078	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	Pr834	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	H1Swj1176	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	H1hmiv	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	Taiw186	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	Chile183	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	HK3283	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H1	HK282	RLKGIAPLQLGKCSNIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H2	Jap30757	KLNGIPPLELGDCSIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H2	Ri5-57	KLNGIPPLELGDCSIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H2	Jap30557	KLNGIPPLELGDCSIAGWILGNPECESLVSKKSWSYIAETPNSENGTCYP
H5	Chpa83	SLNGVKPLILRDCSVAGWLLGNPMCDLFLNAPESYIIVEKNNP INGLCYP
H5	Shaus75	SLNGVKPLILRDCSVAGWLLGNPMCDLFLNAPESYIIVEKNNP INGLCYP
H5	TUIre137883	SLNGVKPLILRDCSVAGWLLGNPMCDLFLNAPESYIIVEKNNP INGLCYP
H5	DUIre11383	SLNGVKPLILRDCSVAGWLLGNPMCDLFLNAPESYIIVEKNNP INGLCYP
H6	Shaus72	KILKKAPLDLKGCTIEGWILGNPQCDLLSVPEWSYIIVEKNNP INGLCYP
H8	Tuont68	NTDLGAPLELRDCKIEAVIYGNPKCDIHLKVNGWSYIVERP.....
H9	Tuwisc166	ATDLGHPLILDCTIEGLIYGNPSCDILLGGKEWSYIIVEKNNP INGLCYP
H10	Duckman53	M-KGRKYKDLGNCHPIGIIIGAPACDLHLTGRWETLI-ERENAI-AYCYP
H10	Minksw84	M-KGRSYKDLGNCHPIGMLIGTPACDLHLTGTWDTLI-ERENAI-AYCYP
H10	Chgern49	M-KGRSYKDLGNCHPIGMLIGTPVCDPHLTGTWDTLI-ERENAI-AYCYP
H12	Dualb6076	GTELGSPLVLDDCSLEGLILGNPKCDLYLNGREWSYIVERP.....
H7	Stvic85	T-KGKKAIDLGCGLLGIITGPPQCDQFLEFTADLII-ERREG-NDVCYP

1000
1000
1000
1000
1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

1000

H7 Chvic85
H7Sealma80
H7 Tuore71
H7 Fpv
H11 Duny78
H11Ternaus75
H11Duukr60
H11Dueng56
H11Dumem76

T-KGKKAIDLGCGLLGIITGPPQCDQFLEFTADLII-ERREG-NDVCYP
T-QGKRPTDLGCGLLGTLLIGPPQCDQFLEFESNLII-ERREG-NDVCYP
T-QGKRPTDLGCGLLGTLLIGPPQCDQFLEFELDLII-ERREG-NNICYP
S-KGKRPTDLGCGLLGTITGPPQCDQFLEFSADLII-ERREG-NDVCYP
SINGKQPI SLGDCSFAGWILGNPMCDLIGKTSWSYIVENQS.....
SINGKQPTSLRDCSFAGWILGNPQC.....
SINGKQPI SLGDCSFAGWILGNPMCDLIGKNSWSYIVENQS.....
SINGKQPI SLGDCSFAGWILGNPMC.....
SIDGKAPISLGDCSFAGWILGNPMCDLIGKTSWSYIVENQS.....

H3 X31
H3 Pc173
H3 Eng4272
H3 Hk10771
H3 Qu770
H3 Eng69
H3 Nt68
H3 X31w68
H3 Vic375
H3 Bk179
H3 Bk279
H3Lenin38580
H3 Eng32177
H3 Mem171
H3 Mem10272
H3 Aichi268
H3 Nt606829c
H3 Tex77
H3 Mem10772
H3 Duukr63
H3 EQ urg63
H3 EQ mia63
H3 EQ tky71
H3 EQ alg72
H3 EQ nm 76
H3 EQ fon79
H3 EQ rom80
H3 EQ san85
H3 EQ tn 86
H3 EQ ky 86
H3 EQ ky 87
H3 DU hk577
H3 DU hk880
H3 DU hk3380
H3 DU hk782
H3 DU hk2182
H3 DU hk985
H3 DU hk1085
H3 SW hk12682
H3 SW hk8178
H4 DU Czk56
H4 GT Aus79
H4 BU Hok77
H4 DU N276
H4 DU Alb76
H4 CH Ala75
H4 RT NJ85
H4 TU Min80

YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP GSGF
YDVPDYASLRSLVASSGTLEFINEGF---TWTGVTQNGG-SNACKRGP DSGF
YDVPDYASLRSLVASSGTLEFINEGF---TWTGVTQNGG-SNACKRGP DSGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTEVTENGG-SNACKRGP DSGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTEVTQNGG-SNACKRGP GSGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP GSGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP GSGF
YDVPDYASLRSLVASSGTLEFINEGF---NWTGVTQNGG-SSACKRGP DSGF
YDVPDYASLRSLVASSGTLEFINEGF---NWTGVTQSGG-SYACKRGS D NSF
YDVPDYASLRSLVASSGTLEFINEGF---NWTGVTQSGG-SYACKRGS D NSF
YDVPDYASLRSLVASSGTLEFINEGF---NWTGVTQNGG-SYACKRGP D NSF
YDVPDYASLR.....-.....RGPGSGF
YDVPDYASLRLLVASSGTLEFINEGF---TLTGVTQNGG-SNACKRGP DSGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP GSGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP DSGF
YDVPDYASLRLLVASSGTLEFINEGF---TLTGVTQNGG-SNACKRGP DSGF
YDIPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SSACKRGP ANGF
YDIPDYASLRSLVASSGTLEFMAEGF---TWTGVTQNGR-SSACRRGSADSF
YDVPDYASLRSLVASSGTLEFMAEGF---TWTGVTQNGG-SSACRRGSADSF
YDVPNYALLRSIVASSGTLEFMAEGF---TWTGVTQNGR-SSACRRGSADSF
YDVPDYASLRSLVASSGTLEFMAEGF---TWTGVTQNGR-SSACRRGSADSF
YDIPDYASLRSLVASSGTLEFTAEGF---TWTGVTQNGR-SGACRRGSADSF
YDIPDYASLRSLVASSGTLEFTAEGF---TWTGVTQNGR-SGACRRGSADSF
YDIPDYASLRSLVASSGTLEFTAEGF---TWTGVTQNGG-SGACRRGSADSF
YDIPDYASLRSLVASSGTLEFTAEGF---TWTGVTQNGR-SGACRRGSADSF
YDIPDYASLRSLVASSGTLEFTAEGF---TWTGVTQNGR-SGACRRGSADSF
YDIPDYASLRSLVASSGTLEFTAEGF---TWTGVTQNGR-SGACRRGSADSF
YDIPDYASLRSLVASSGTLEFTAEGF---TWTGVTQNGR-SGACRRGSADSF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP ASGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP ASGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP ASGF
YDVSDYASLRSLVASSGTLEFLTEGF---TWTGVTQNGG-SNACKRGP ASGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP NSGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP NSGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SSACKRGP ASGF
YDVPDYASLRSLVASSGTLEFITEGF---TWTGVTQNGG-SNACKRGP ANGF
FDVPEYQSLRSILANNGKFEFIAEEF---QWNTVKQNGK-SGACKRANV NDF
FDVPEYQSLRSILANNGKFEFIAEEF---QWNTVKQNGK-SGACKRANV NDF
FDVDPDYQSLRSILANNGKFEFIAEEF---QWNTVIQNGK-SSACKRANV NDF
FDVDPDYQSLRSILANNGKFEFIAEEF---QWSTVKQNGK-SGACKRANINDF
FDVDPDYQSLRSILANNGKFEFIAEEF---QWNTVKQNGK-SGACKRANV NDF
FDVDPDYQSLRSILANNGKFEFIVEKF---QWNTVKQNGK-SGACKRANENDF
FDVDPDYQSLRSILANNGKFEFIAEEF---QWNTVKQNGK-SGACKRANV NDF
FDVDPDYQSLRSILANNGKFEFIAEEF---QWNTVKQNGK-SGACKRANV NDF

1. The first part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the name of the office to which the person has been appointed.

2. The second part of the document is a list of the names of the persons who have been appointed to the various offices of the city government. The names are listed in alphabetical order, and each name is followed by the name of the office to which the person has been appointed.

H4 SE Mas82	FDVPDYQSLRSILANNGKFEFIAEEF---QWSTVKQNGD-SGTCKRGNVNGF
H1 Khab77	GYFADYEELREQLSSVSSFERFEIFPKERSWPKHNVTRGVTASC SHKGKSSF
H1 Kiev5979	GYFADYEELREQLSSVSSFERFEIFPKERSWPKHNVTRGVTASC SHKGKSSF
H1 Wsn33	GDFIDYEELREQLSSVSSLERFEIFPKESSWPNHFTN-GVTVSCSHRGKSSF
H1 Pr834	GDFIDYEELREQLSSVSSFERFEIFPKESSWPNHNTTKGVTAACSHAGKSSF
H1 Swnj1176	GDFINYEELREQLSSVSSFERFEIFPKTSSWPNHETNRGVTAAACPYAGANSF
H1 Hlhmiv	GDFIDYEELREQLSSVSSFERFEIFPKESSWPNHNTN-GVTAACPHEGKSSF
H1 Taiw186	GYFADYEELREQLSSVSSFERFEIFPKTSSWPNHNTVTKGVTAACSHKGKSSF
H1 Chile183	GYFADYEELREQLSSVSSFERFEIFPKTSSWPKHNVTKGVTAACSHKGKSSF
H1 HK3283	GYFADYEELREQLSSVSSFERFEIFPKTSSWPKHNVTKGVTAACSHKGKSSF
H1 HK282	GYFADYEELREQLSSVSSFERFEIFPKTSSWPKHNVTKGVTAACSHKGKSSF
H2 Jap30757	GSFNDYEELKHLSSVKHFEKVKILPKDR-WTQHTTTGG-SRACAVSGNP SF
H2 Ri5-57	G.....
H2 Jap30557	GSFNDYEELKHLSSVKHFEKVKILPKDR-WTQHTTTGG-SRACAVSGNP SF
H5 Chpa83	GDFNDYEELKHLVSSSTNLF EKIRIIPRNS-WTNHDASSGVSSACPHLGRSSF
H5 Shaus75	G.....
H5 TUIrel137883	GDFNDYEELKHLSSCTKHFEKIRIIPRDS-WPNHDASLGVS SACPYNGRSSF
H5 DUIrel1383	GDFNDYEELKHLSSCTKHFEKIRIIPRDS-WPNHDASLGVS SACPYNGRSSF
H10Minksw84	GTTINEGALRQKIMESGGISKSTSGF---AYGSSINSAGTTKACMRNGGDSF
H10Chgern49	GATINEEALRQKIMESGGISKSTSGF---TYGSSITSAGTTKACMRNGGDSF
H7 Stvic85	GK FVNEEALRQILRESGGINKETTGF---TYSGIRTN-GVTSACRRSG-SSF
H7 Chvic85	GK FVNEEALRQILRESGGINKETTGF---TYSGIRTN-GVTSACRRSG-SSF
H7 Sealma80	GKFTNEESLRQILRGSGGVDKESMGF---TYSGIRTN-GTTSACRRSG-SSF
H7 Tuore71	G.....
H7 Fpv	GK FVNEEALRQILRGSGGIDKETMGF---TYSGIRTN-GTTSACRRSG-SSF
H3 X31	FSRLNWLT KSGS--TYPVLNVTMPNNDNFDKLYIWGIHHPSTNQEQTSLY
H3 Pc173	FSRLNWLYKSGS--AYPVLNVTMPNNDNFDKLYIWGVHHPSTDQEQT NLY
H3 Eng4272	FSRLNWLYKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTSLY
H3 Hk10771	FSRLNWLT KSGR--TYPVLNVTMPNNDNFDKLYIWGVHHPSTDQEQT NLY
H3 Qu770	FSRLNWLT KSGS--TYPVLNVTMPNNDNFDKLYIWGIHHPSTNQEQTSLY
H3 Eng69	FSRLNWLT KSGS--TYPVLNVTMPNNDNFDKLYIWGIHHPSTNQEQTSLY
H3 Nt68	FSRLNWLT KSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTSLY
H3 X31w68	FSRLAWLT KSES--TYPVLNVTMPNNDNFDKLYIWGVHHPSTDQEQT NLY
H3 Vic375	FSRLNWLYKSGS--TYPVQNVTMPNNDNSDKLYIWGVHHPSTDKEQT NLY
H3 Bk179	FSRLNWLYESES--KYPVLNVTMPNNGNFDKLYIWGVHHPSTDKEQT NLY
H3 Bk279	FSRLNWLYESES--KYPVLNVTMPNNGNFDKLYIWGVHHPSTYKEQT NLY
H3 Lenin38580	FSRLNWLYESES--KYPVLNVTMPNNGNFDKLYIWGVHHPSTDKEQT NLY
H3 Eng32177	FSRLNWLYKSES--TYPVLNVTMPNNDNFDKLYIWGVHHPSTDKEQT KLY
H3 Mem171	FSRLNWLT KS-S--TYPVLNVTMPNNDNFDKLYIW-VHHPSTDQEQT NLY
H3 Mem10272	FSRLNWLYKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTDQEQT NLY
H3 Aichi268	FSRLNWLT KSGS--TYPVLNVTMPNNDNFDKLYIWGIHHPSTNQEQTSLY
H3 Nt606829c	FSRLNWLT KSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTSLY
H3 Tex77	FSRLNWLYKSES--TYPVLNVTMPNNGNFDKLYIWGVHHPSTDKEQT NLY
H3 Mem10772	FSRLNWLYKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTDQEQT NLY
H3 Duukr63	FSRLNWLT KSES--AYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTSLY
H3 EQ urg63	FSRLNWLT KSG-N-SYPTLNVTMPNNDNFDKLYIWGIHHPSTNNEQT KLY
H3 EQ mia63	FSRLNWLT KSE-S-SYPTLNVTMPNNDNFDKLYIWGIHHPSTNNEQT KLY
H3 EQ tky71	FSRLNWLT KSG-S-SYSTLNVTMPNNDNFDKLYVWGIHHPSTNDEQT KLY
H3 EQ alg72	FSRLNWLT KSE-S-SYSTLNVTMPNNDNFDKLYIWGIHHPSTNNEQT KLY
H3 EQ nm 76	FSRLNWLT KSG-D-SYPTLNVTMPNNDNFDKLYIWGIHHPSTNNEQT KLY
H3 EQ fon79	FSRLNWLT KSG-N-SYPTLNVTMPNNDNFDKLYIWGIHHPSTNNEQT KLY
H3 EQ rom80	FSRLNWLT KSG-N-SYPTLNVTMPNNDNFDKLYIWGIHHPSSNNEQT KLY
H3 EQ san85	FSRLNWLT KSG-N-SYPTLNVTMPNNDNFDKLYIWGIHHPSSNNEQT KLY
H3 EQ tn 86	FSRLNWLT KSG-N-SYPTLNVTMPNNDNFDKLYIWGIHHPSSNNEQT KLY
H3 EQ ky 86	FSRLNWLT KSG-N-SYPTLNVTMPNNDNFDKLYIWGIHHPSSNNEQT KLY
H3 EQ ky 87	FSRLNWLT KSG-N-SYPTLNVTMPNNDNFDKLYIWGIHHPSSNNEQT KLY
H3 DU hk577	FSRLNWLT KSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTDQEQT NLY

H3 DU hk880	FSRLNWLTKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTNLY
H3 DU hk3380	FSRLNWLTKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTNLY
H3 DU hk782	FSRLNWLTKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTNLY
H3 DU hk2182	FSRLNWLTKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTDLY
H3 DU hk985	FSRLNWLTKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTNLY
H3 DU hk1085	FSRLNWLTKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTNLY
H3 SW hk12682	FSRLNWLTKSGS--TYPVLNVTMPNNDNFDKLYIWGVHHPSTNQEQTNLY
H3 SW hk8178	FSRLNWLTKSGS--TYPVLNVTMPNNDNSDKLYIWGVHHPSTNQEQTNLY
H4 DU Czk56	FNRLNWLKSDGN--AYPLQNLTKINNGDYARLYIWGVHHPSTDTEQTNLY
H4 GT Aus79	FNRLNWLKSDGN--AYPLQNLTKINNGDYARLYIWGVHHPSTDTEQTNLY
H4 BU Hok77	FNRLNWLKSTGN--AYPLQNLTKVNNGDYARLYIWGVHHPSTDTEQTNLY
H4 DU NZ76	FNRLNWLKSDGN--AYPLQNLTKINNGDYARLYIWGVHHPSTDTEQTNLY
H4 DU Alb76	FNRLNWLKSDGD--AYPLQNLTKVNNGDYARLYIWGVHHPSTDTEQTNLY
H4 CH Ala75	FTNLNWLTKSDGN--AYPLQNLTKVNNGDYARLYIWGVHHPSTDTEQTNLY
H4 RT NJ85	FRRLNWLTKSDGN--AYPLQNLTKVNNGDYARLYIWGVHHPSTDTEQTNLY
H4 TU Min80	FNRLNWLTKSDGN--AYPLQNLTKVNNGDYARLYIWGVHHPSTDTEQTNLY
H4 SE Mas82	FRQLNWLTKSNGD--AYPLQNLTKANDGDYARLYIWGVHHPSTDTEQTNLY
H1 Khab77	YRNLLWLTEKNG--SYPNLSKSYVNNKEKEVLVLWGVHHP SNIEDQKTIY
H1 Kiev5979	YRNLLWLTEKNG--SYPNLSKSYVNNKEKEVLVLWGVHHP SNIEDQKTIY
H1 Wen33	YRNLLWLTKKGD--SYPKLTNSYVNNKGKEVLVLWGVHHPSSSDEQQSLY
H1 Pr834	YRNLLWLTEKEG--SYPKLKNSYVNNKGKEVLVLWGIHHP SNSKDQQNIY
H1 Swj1176	YRNLIWLVKKEN--SYPKLKNSYVNNKGKEVLVLWGIHHPPTSTDQQSLY
H1 Hihmiv	YRNLLWLTEKEG--SYPKLKNSYVNNKGKEVLVLWGIHHP SNSKEQQNLY
H1 Taiw186	YRNLLWLTEKNG--SYPNLSKSYVNNKEKEVLVLWGVHHP SNIGDQRAIY
H1 Chile183	YRNLLWLTEKNG--SYPNLSKSYVNNKEKEVLVLWGVHHP SNIEDQKTIY
H1 HK3283	YRNLLWLTEKNG--SYPNLSKSYVNNKEKEVLVLWGVHHP SNIEDQKTIY
H1 HK282	YRNLLWLTEKNG--SYPNLSKSYVNNKEKEVLVLWGVHHP SNIKDQKTIY
H2 Jap30757	FRNMVWLTKEGS--DYPVAKGSYNNTSGEQMLI IWGVHHP IDETEQRTLY
H2 Jap30557	FRNMVWLTKEGS--DYPVAKGSYNNTSGEQMLI IWGVHHP IDETEQRTLY
H5 Chpa83	FRNVVWLIKKN--VYPTIKRTYNNTNVEDLLILWGIHHPNDAAEQAKLY
H5 TUIre137883	FRNVVWLIKKN--AYPTIKRSYNTNKEDLLILWGIHHPNDAAEQTKLY
H5 DUIre11383	FRNVVWLIKKN--AYPTIKRSYNTNKEDLLILWGIHHPNDAAEQTKLY
H10Minksw84	YAEVKWLVSCKDKQNFQTTNTYRNTDTAEHLI IWGIHHP SSTQEKNDLY
H10Chgern49	YAEVKWLVSCKTKQNFQTTNTYRNTDTAEHLI IWGIHHP SSTQEKNDLY
H7 Stvic85	YAEMKWLLSNTDNAAFQMTKSYKNTRNEPALIVWGIHHS GSATEQTKLY
H7 Chvic85	YAEMKWLLSNTDNAAFQMTKSYKNTRNEPALIVWGIHHS GSATEQTKLY
H7 Sealma80	YAEMKWLLSNTDNAAFQMTKSYRNP RNKPALIVWGIHHS GSSTTEQTRLY
H7 Fpv	YAEMEWLLSNTDNAAFQMTKSYKNTRRESALIVWGIHHS GSSTTEQTKLY
H3 X31	VQASGRVTVSTRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDVLVINSNG
H3 Pc173	VQTSGRVTVSTKRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDVLVINSNG
H3 Eng4272	VQASGRVTVSTKGSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDILVINSNG
H3 Hk10771	VQTSGRVTVSTRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDVLVINSNG
H3 Qu770	VQASGRVTVSTRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDVLVINSNG
H3 Eng69	VQASGRVTVSTRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDVLVINSNG
H3 Nt68	VQASGRVTVSTRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDVLVINSNG
H3 X31w68	VQASGRVTVSTRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDVLVINSNG
H3 Vic375	VQASGKVTVSTKRSQQTIIIPNVGSRPWVRLSSRISIYWTIVKPGDILVINSNG
H3 Bk179	VRASGRVTVSTKRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDILLINSNG
H3 Bk279	VRASGRVTVSTKRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDILLINSNG
H3 Lenin38580	VRASGRVTVSTKRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDILLINSNG
H3 Eng32177	VQASGRVTVSTKRSQQTIIIPNVGSRPWVRLSSRISIYWTIVKPGDILLINSNG
H3 Mem171	VQASGRVTVSTRSQQTIIIPNIGSRPWVRLSSR.....
H3 Mem10272	VQASGRVTVSTKRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDILVINSNG
H3 Aichi268	VQASGRVTVSTRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDVLVINSNG
H3 Nt606829c	VQASGRVTVSTRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDVLVINSNG
H3 Tex77	VQASGRVTVSTKRSQQTIIIPNVGSRPWVRLSSGISIYWTIVKPGDILLINSNG
H3 Mem10772	VQASGRVTVSTKRSQQTIIIPNIGSRPWVRLSSRISIYWTIVKPGDILVINSNG

1000
1000
1000
1000
1000

1000
1000
1000
1000
1000

1000
1000
1000
1000
1000

H3	Duukr63	VQASGRVTVSTRRSQQTIIIPNIGSRPWVRGQPGRISIIYWTIVKPGDVLVINSNG
H3	EQ urg63	VQASGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	EQ mia63	VQASGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	EQ tky71	VQASGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	EQ alg72	VQASGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	EQ nm 76	VQESGRVTVSTKRSQQTIIIPNIGSRPWVRGQAGRISIIYWTIVKPGDVLVINSNG
H3	EQ fon79	VQELGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDILVINSNG
H3	EQ rom80	VQELGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDILVINSNG
H3	EQ san85	VIESGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDILVINSNG
H3	EQ tn 86	VIESGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDILVINSNG
H3	EQ ky 86	VIESGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDILVINSNG
H3	EQ ky 87	VIESGRVTVSTKRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDILVINSNG
H3	DU hk577	VQASGRVTVSTRRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	DU hk880	VQASGRVTVSTRRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	DU hk3380	VQASGGVTVSTRRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	DU hk782	VQASGRVTVSTRRSQQTIIIPNIGSRPWVRGQSSRISIIYWTIVKPGDVLVINSNG
H3	DU hk2182	VQASGRVTVSTRRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	DU hk985	VQASGGVTVSTRRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	DU hk1085	VQASGRVTVSTRRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	SW hk12682	VQASGRVTVSTRRSQQTIIIPNIGSRPWVRGQSGRISIIYWTIVKPGDVLVINSNG
H3	SW hk8178	VQASGRVTVSTKRSQQTIIIPNAGSRPWVRGLSSRISIIYWTIVKPGDILVINSNG
H4	DU Czk56	KNNPGSVTVSTKTSQTSVVPNIGSRPLVRGQSGRISFYWTIVEPGLDIFNTIG
H4	GT Aus79	KNNPGRVTVSTKTSQTSVVPNIGSRPLVRGQSGRISFYWTIVEPGLDIFNTIG
H4	BU Hok77	KNNPGRVTVSTKTSQTSVVPNIGSRPLVRGQSGRISFYWTIVEPGLDIFNTIG
H4	DU NZ76	KNNPGRVTVSTKTSQTSVVPNIGSRPLVRGQSGRISFYWTIVEPGLDIFNTIG
H4	DU Alb76	KNNPGRVTVSTKTSQTSVVPNIGSRPWVRGQSGRISFYWTIVEPGLDIFNTIG
H4	CH Ala75	ENNPGRVTVSTKTSQTSVVPNIGSRPWVRGQSGRISFYWTIVEPGLDIFNTIG
H4	RT NJ85	KNNPGRVTVSTKTSQTSVVPNIGSRPLVRGQSGRISFYWTIVEPGLDIFNTIG
H4	TU Min80	ENNPGRVTVSTKTSQTSVVPNIGSRPWVRGQSGRISFYWTIVEPGLDIFNTIG
H4	SE Mas82	KNNPGRVTVSTKTSQTSVVPNIGSRPWVRGQSGRISFYWTIVEPGLDIFNTIG
H1	Khab77	RKENAYVSVVSSSNYNRRFTPEIAERPKVRGQAGRINYYWTLLEPGDTIIFEANG
H1	H1Kiev5979	RKENAYVSVVSSSNYNRRFTPEIAERPKVRGQAGRINYYWTLLEPGDTIIFEANG
H1	Wsn33	SNGNAYVSVASSSNYNRRFTPEIAARPKVKDQHGRMNYWTLLEPGDTIIFEATG
H1	Pr834	QENAYVSVVTSNYNRRFTPEIAERPKVRDQAGRMNYWTLLEPGDTIIFEANG
H1	H1Swj1176	QNADAYVVFVGSSSKYNRKFKPEIAARPKVRGQAGRMSYYWTLLEPGDTIIFEATG
H1	H1hmiv	QENAYVSVVTSNYNRRFTPEIAERPKVRDQAGRMNYWTLLEPGDTIIFEANG
H1	Taiw186	HTENAYVSVVSSHYNRRFTPEIAERPKVRGQAGRINYYWTLLEPGDTIIFEANG
H1	Chile183	RKENAYVSVVSSHYNRRFTPEIAERPKVRNQEGRINYYWTLLEPGDTIIFEANG
H1	HK3283	RKENAYVSVVSSHYNRRFTPEIAERPKVRGQAGRINYYWTLLEPGDTIIFEANG
H1	HK282	RKENAYVSVVSSHYNRRFTPEIAERPKVRDQAGRINYYWTLLEPGDTIIFEANG
H2	Jap30757	QNVGTYSVGTSTLNKRSTPGIATRPKVNGQGRMEFSWTLLDMWDITINFESTG
H2	Jap30557	QNVGTYSVGTSTLNKRSTPGIATRPKVNGQGRMEFSWTLLDMWDITINFESTG
H5	Chpa83	QNLNAYVSV-TSTLNQRSIPKIATRPKVNGQSGRMEFFWTILKPNDTISFESTG
H5	TUIrel137883	QNPTTYVSVGTSTLNQRSIPKIATRPKLNGQSGRMEFFWTILKPNDTISFESNG
H5	DUIrel11383	QNPTTYVSVGTSTLNQRSIPKIATRPKLNGQSGRMEFFWTILKPNDTISFESNG
H10	Minksw84	GTQSLSISVGSSTYQNNFVPPVVRARPQVNGQSGRIDFHWTLVQPGDNITFSHNG
H10	Chgern49	GTQSLSISVESSTYQNNFVPPVVGARPQVNGQSGRIDFHWTLVQPGDNITFSHNG
H7	Stvic85	GSGNKLITVGSSSNYQQSFVSPGARPQVNGQSGRIDFHWLILNPNDTVTFSFNG
H7	Chvic85	GSGNKLITVGSSSNYQQSFVSPGARPQVNGQSGRIDFHWLILNPNDTVTFSFNG
H7	Sealma80	GSGNKLITVGSSSKYQSFVSPGARPQVNGQSGRIDFHWLILNPNDTVTFSFNG
H7	Fpv	GSGNKLITVGSSSKYHQSFPVSPGTRPQINGQSGRIDFHWLILNPNDTVTFSFNG
H3	X31	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPFQON
H3	Pc173	NLIAPRGYFKMRTGKSS-IMRSDAPIGTCISECITPNGSIPNDKPFQON
H3	Eng4272	NLIAPRGYFKMRTGKSS-IMRSDAPIGTCISECITPNGSIPNDKPFQON
H3	Hk10771	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNG-IPND-PFQON
H3	Qu770	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPFQON
H3	Eng69	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPFQON

1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 2679, 26

•

7

H3	Nt68	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPPFQ
H3	X31w68	DLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPPFQ
H3	Vic375	NLIAPRGYFKMRTGKSS-IMRSDAPIGTCSSECITPNGSIPNDKPPFQ
H3	Bk179	NLIAPRGYFKIRTGKSS-IMRSDAPIGTCSSECITPNGSIPNDKPPFQ
H3	Bk279	NLIAPRGYFKIRTGKSS-IMRSDAPIGTCSSECITPNGSIPNDKPPFQ
H3	Lenin38580	NLIAPRGYFKIRTGKSS-IMRSDAPIGTCSSECITPNGSIPNDKPPFQ
H3	Eng32177	NLIAPRGYFKIRTGKSS-IMRSDAPIGTCSSECITPNGSIPNDKPPFQ
H3	Mem171GYFKMRTGKSS-IMR.....
H3	Mem10272	NLIAPRGYFKMRTGKSS-IMRSDAPIGTCSSECITPNGSIPNDKPPFQ
H3	Aichi268	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPPFQ
H3	Nt606829c	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPPFQ
H3	Tex77	NLIAPRGYFKIRTGKSS-IMRSDAPIGTCSSECITPNGSIPNDKPPFQ
H3	Mem10772	NLIAPRGYFKMRTGKSS-IMRSDAPIGTCSSECITPNGSIPNDKPPFQ
H3	Duukr63	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPPFQ
H3	EQ urg63	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCVSECITPNGSIPNNKPPFQ
H3	EQ mia63	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	EQ tky71	NLIAPRGYFKIRAGKSS-IMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	EQ alg72	NLIAPRGYFKMRAGKSS-IMRSDAPIDTCVFECITPNGSIPNDKPPFQ
H3	EQ nm 76	NLVAPRGYFKMRTEKSS-IMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	EQ fon79	NLVAPRGYFKMRTGKSS-IMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	EQ rom80	NLVAPRGYFKMRTGKSS-IMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	EQ san85	NLVAPRGYFKMRTGKSS-VMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	EQ tn 86	NLVAPRGYFKLRTGKSS-VMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	EQ ky 86	NLVAPRGYFKLRTGKSS-VMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	EQ ky 87	NLVAPRGYFKLRTGKSS-VMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	DU hk577	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPPFQ
H3	DU hk880	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	DU hk3380	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPPFQ
H3	DU hk782	NLIAPRGYFKMRTGKSS-VMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	DU hk2182	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECVTPNGSIPNDKPPFQ
H3	DU hk985	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPPFQ
H3	DU hk1085	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCISECITPNGSIPNDKPPFQ
H3	SW hk12682	NLIAPRGYFKMRTGKSS-IMRSDAPIDTCVSECITPNGSIPNDKPPFQ
H3	SW hk8178	NLIAPRGYFKMRTGKSS-IMRSDAPIGTCSSECITPNGSIPNDKPPFQ
H4	DU Czk56	NLIAPRGYKLNQKSTILNTAIPIGSCVSKCHTDKGSLSSTTKPPFQ
H4	GT Aus79	NLIASRGHYKVNNQKSTILNTAIPIGSCVSKCHTDKGSLSSTTKPPFQ
H4	BU Hok77	NLIAPRGHYKLNQKSTILNTP IPIGSCVSKCHTDKGSVSTTNPPFQ
H4	DU NZ76	NLIAPRGHYKLNQKSTILNTAVPIGSCVSRCHTDKGSLSRTKPPQA
H4	DU Alb76	NLIAPRGHYKLNSQKSTILNTAVPIGSCVSKCHTDGRGSIITTKPPFQ
H4	CH Ala75	NLIAPRGHYKLNSQKSTILNTAVPIGSCVSKCHTDGRGSIITTKPPFQ
H4	RT NJ85	NLIAPRGHYKLNSQKSTILNTAVPIGSCVSKCHTNRGSIITTKPPFQ
H4	TU Min80	NLIAPRGHYKLNSQKSTILNTAVPIGSCVSKCHTDGRGSIITTKPPFQ
H4	SE Mas82	NLIAPRGHYKLNSQKSTILNTAVPIGSCTSKCHTDGRGSIITTKPPFQ
H1	Khab77	NLIAPWHAFALNRGFGSGIITSNASMDECDTKCQTPQGAINSSLPPFQ
H1	Kiev5979	NLIAPWYAFALSRGFGSGIITSNASMDECDTKCQTPQGAINSSLPPFQ
H1	Wsn33	NLIAPWYAFALSRGFGSGIITSNASMHECNTKCQTPQGSINSLPPFQ
H1	Pr834	NLIAPRYAFALSRGFGSGIITSNASMHECNTKCQTPQGAINSSLPPFQ
H1	Swj1176	NLVVPRYAFAMNRGSGSGIIWDAPVHDCNTKCQTPKGAINSSLPPFQ
H1	H1hmiv	NLIAPMYAFALSRGFGSGIITSNASMHECNTKCQTPLGAINSSLPPYQ
H1	Taiw186	NLIAPWYAFALSRGFGSGIITSNASMDECDAKCQTPQGAINSSLPPFQ
H1	Chile183	NLIAPWYAFALSRGFGSGIITSNASMDECDAKCQTPQGAINSSLPPFQ
H1	HK3283	NLIAPWYAFALSRGFGSGIITSNASMDECDAKCQTPQGAINSSLPPFQ
H1	HK282	NLIAPWYAFALSRGFGSGIITSNASMDECDAKCQTPQGAINSSLPPFQ
H2	Jap30757	NLIAPEYGFKISKRGSSGIMKTEGTLENCETKCQTPLGAINSSLPPFH
H2	Jap30557	NLIAPEYGFKISKRGSSGIMKTEGTLENCETKCQTPLGAINSSLPPFH
H5	Chpa83	NFIAPEYAYKIVKKGDSAIMRSELEYGNCDTKCQTPLVAINSSMPFH
H5	TUIrel137883	NFIAPEYAYKIVKKGDSAIMKSGLEYGNCNTKCQTPIGAINSSMPFH
H5	DUIrel11383	NFIAPEYAYKIVKKGDSAIMKSGLEYGNCNTKCQTPIGAINSSMPFH

1

GRIAPSRVSKLVGRGL-GIQSEASIDNGCESKCFWRGGSINTKLPFQN
GLIAPSRVSKLTGRDL-GIQSEALIDNSCESKCFWRGGSINTKLPFQN
AFVAPDRVVSFF-KGKSMGIQSEVPVDTNCEGECYHNGGTITSNLPFQN
AFVAPDRVVSFF-KGKSMGIQSEVPVDTNCEGECYHNGGTITSNLPFQN
AFIAPNRSFRL-RGESLGVQSDVPLDSCNGGDCYHSGGTIVSSLPFQN
AFIAPNRSFRL-RGKSMGIQSDVQVDANCEGECYHSGGTITSRLPFQN

[illegible]

H1	Wsn33	IHPVTIGECPKYVRSTKLRMTGLRNIPSIQY
H1	Pr834	IHPVTIGECPKYVRSAKLRMTGLRNIPSIQS
H1	Swnj1176	IHPVTIGECPKYVKSTKLRMATGLRNVP SIQS
H1	Hlhmiv	IHPVTIGECPKYVRSAKLRMTGLRNIPSIQS
H1	Taiw186	VHPVTIGECPKYVRSTKLRMTGLRNIPSIQS
H1	Chile183	VHPVTIGECPKYVRSTKLRMTGLRNIPSIQS
H1	HK3283	VHPVTIGECPKYVRSTKLRMTGLRNIPSIQS
H1	HK282	VHPVTIGECPKYVRSTKLRMTGLRNIPSIQS
H2	Jap30757	VHPLTIGECPKYVKSEKLVLATGLRNVPQIES
H2	Jap30557	VHPLTIGECPKYVKSEKLVLATGLRNVPQIES
H5	Chpa83	VHPLTIGECPKYVKSDKLVLATGMRNV PQKKK
H5	TUIre137883	IHPLTIGECPKYVKSDRLVLATGLRNTPQRKRKKR
H5	DUIre11383	IHPLTIGECPKYVKSDRLVLATGLRNTPQRKRKKR
H10	Minksw84	LSPRTVGQCPKYVNKKSIMLATGMRNVPEIMQG
H10	Chgern49	LSPRTVGQCPKYVNQRSLLLATGMRNVPEVVQG
H7	Stvic85	VNSRAVGKCPRYVKQKSLLLATGMKNVPEIPKKREKR
H7	Chvic85	VNSRAVGKCPRYVKQKSLLLATGMKNVPEIPKKREKR
H7	Sealma80	INSRTVGKCPRYVKQPSLLLATGMRNV PENPKT
H7	Fpv	INSRAVGKCPRYVKQESLLLATGMKNVPEPSKKREKR

HA2

H3	X31	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Nt68	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Vic375	GIFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Bk179	GIFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Bk279	GIFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Lenin38580	GIFGAIAGFIENGWEGMVDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Eng32177	GIFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Mem10272	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Aichi268	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Nt606829c	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Mem10772	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	Duukr63	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQINR
H3	EQ urg63	GIFGAIAGFIENGWEGMIDGWYGFYQNSEGTGQAADLKSTQAAIDQING
H3	EQ mia63	GIFGAIAGFIENGWEGMVDGWYGFYQNSEGTGQAADLKSTQAAIDQING
H3	EQ tky71	GIFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	EQ alg72	GIFGAIAGFIENGWEGMVDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	EQ nm 76	GIFGAIAGFIENGWEGMVDGWYGFYQNSEGTGQAADLKSTQAAIDQING
H3	EQ fon79	GIFGAIAGFIENGWEGMVDGWYGFYQNSEGTGQAADLKSTQAAIDQING
H3	EQ rom80	GIFGAIAGFIENGWEGMVDGWYGFYQNSEGTGQAADLKSTQAAIDQING
H3	EQ san85	GIFGAIAGFIENGWEGMVDGWYGFYQNSEGTGQAADLKSTQAAIDQING
H3	EQ tn 86	GIFGAIAGFIENGWEGMVDGWYGFYQNSEGTGQAADLKSTQAAIDQING
H3	EQ ky 86	GIFGAIAGFIENGWEGMVDGWYGFYQNSEGTGQAADLKSTQAAIDQING
H3	EQ ky 87	GIFGAIAGFIENGWEGMVDGWYGFYQNSEGTGQAADLKSTQAAIDQING
H3	DU hk577	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	DU hk880	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	DU hk3380	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	DU hk782	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	DU hk2182	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	DU hk985	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	DU hk1085	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	SW hk12682	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQING
H3	SW hk8178	GLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLKSTQAAIDLING
H4	DU Czk56	GLFGAIAGFIENGWQGLIDGWYGFRHQNAEGTGTAADLKSTQAAIDQING
H4	GT Aus79	GLFGAIAGFIENGWQGLIDGWYGFRHQNAEGTGTAADLKSTQAAIDQING

H4 DU Czk56
 H4 GT Aus79
 H4 BU Hok77
 H4 DU NZ76
 H4 DU Alb76
 H4 CH Ala75
 H4 RT NJ85
 H4 TU Min80
 H4 SE Mas82
 H1 Khab77
 H1Kiev5979
 H1 Wsn33
 H1 Pr834
 H1Swnj1176
 H1 Hlhmiv
 H2 Jap30757
 H2 Jap30557
 H5 Chpa83
 H5 TUIrel137883
 H5 DUIrel1383
 H10Minksw84
 H10Chgern49
 H7 Stvic85
 H7 Chvic85
 H7Sealma80
 H7 Fpv

KLNRLIEKTNDKYHQIEKEFEQVEGRIQDLEKYVEDTKIDLWSYNAELLV
 KLNRLIEKTNDKYHQIEKEFEQVEGRIQDLKKYVEDTKIDLWSYNAELLV
 KLNRLIEKTNEKYHQIEKEFNKIEGRVQDLEKYVEDTKIDLWSYNAELLV
 KLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDLWSYNAEFLV
 KLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDLWSYNAEFLV
 KLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDLWSYNAELLV
 KLNRLIEKTNEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDLWSYNAELLV
 KLNRLIEKTSEKYHQIEKEFEQVEGRIQDLEKYVEDTKIDLWSYNAELLV
 KVNSVIEKMNTQFTAVGKDFDELEKRMENLNKKVDDGFLDIWTYNAELLV
 KVNSVIEKMNTQFTAVGKDFNKLKRMENLNKKVDDGFLDIWTYNAELLV
 KVNSVIEKMNTQFTAVGKEFNKLEKRMENLNKKVDDGFLDIWTYNAELLV
 KVNSVIEKMNTQFTAVGKEFNHLEKRIENLNKKVDDGFLDIWTYNAELLV
 KVNSVIEKMNTQFEAVGKEFGNLERLENLNKRMEDGFLDVWTYNAELLV
 KVNSVIEKMNTQFEAVGKEFGNLERLENLNKRMEDGFLDVWTYNAELLV
 KVNSIIDKMNTQFEAVGREFNNLERRIENLNKNLEDGFLDVWTYNAELLV
 KVNSIIDKMNTQFEAVGKEFNNLERRIENLNKILEDGFLDVWTYNAELLV
 KVNSIIDKMNTQFEAVGKEFNNLERRIENLNKILEDGFLDVWTYNAELLV
 KLNRLIEKTNTFESIESEFSEIEHQIGNVINWTKDSITDIWTYQAEELV
 KLNRLIEKTNTFESIESEFSETEHQIGNVINWTKDSITDIWTYNAELLV
 KLNRLIEKTNQFELIDNEFTEVEKQIGNVINWTRDSITEVWSYNADLLV
 KLNRLIEKTNQFELIDNEFTEVEKQIGNVINWTRDSITEVWSYNADLLV
 KLNRLIDKTNQFELIDNEFNEIEQQIGNVINWTRDSMTEVWSYNAELLV
 KLNRLIEKTNQFELIDNEFTEVEKQIGNLINWTKDFITEVWSYNAELLV

H3 X31
 H3 Nt68
 H3 Vic375
 H3 Bk179
 H3 Bk279
 H3Lenin38580
 H3 Eng32177
 H3 Mem10272
 H3 Aichi268
 H3 Nt606829c
 H3 Mem10772
 H3 Duukr63
 H3 EQ urg63
 H3 EQ mia63
 H3 EQ tky71
 H3 EQ alg72
 H3 EQ nm 76
 H3 EQ fon79
 H3 EQ rom80
 H3 EQ san85
 H3 EQ tn 86
 H3 EQ ky 86
 H3 EQ ky 87
 H3 DU hk577
 H3 DU hk880
 H3 DU hk3380
 H3 DU hk782
 H3 DU hk2182
 H3 DU hk985
 H3 DU hk1085
 H3 SW hk12682

ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIE
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIE
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIG
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIG
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIG
 ALENQHTIELTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIG
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIG
 ALGNQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIG
 ALENQHTIDLTDSEMNKLFETRRLRENAEEMGNGCFKIYHKCDNACIE
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIE
 ALGNQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIG
 ALENQHTIDLADSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIE
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGNGCFKIYHKCDNACIE
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGNGCFKIYHKCDNACIE
 TLENQHTIDLDAEMNKLFEKTRRLRENAEDIGNGCFKIYHKCDNACIE
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGNGCFKIYHKCDNACIE
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGGGCFKIYHKCDNACIG
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGGGCFKIYHKCDNACIG
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGGGCFKIYHKCDNACIG
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGGGCFKIYHKCDNACIG
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGGGCFKIYHKCDNACIG
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGGGCFKIYHKCDNACIG
 ALENQHTIDLDAEMNKLFEKTRRLRENAEDMGGGCFRIYHKCDNACIG
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACVE
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIE
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIE
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIE
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNACIE
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNVCIE
 ALENQHTIDLTDSEMNKLFETRRLRENAEDMGNGCFKIYHKCDNVCIE
 ALENQHTIDLTDSEMNKLFETRRLRENAEEMGNGCFKIYHKCDNACIE

H3	SW	hk8178	ALENQHTIDLTDSEMKNLFETRRLRENAEEMGNGCFKIYHKCDNACIE
H4	DU	Czk56	ALENQHTIDVTDSEMKNLFEVRRQLRENAEDKGNGCFEIFHKCDYNICIE
H4	GT	Aus79	ALENQHTIDVTDSEMKNLFEVRRQLRENAEDKGNGCFEIFHKCDNNCIE
H4	BU	Hok77	ALENQHTIDVTDSEMKNLFEVRRQLRENAEDQGNGCFEIFHKCDNNCIE
H4	DU	NZ76	ALENQHTIDVADSEMKNLFEVRRQLRENAEDKGNGCFEIFHKCDNNCIE
H4	DU	Alb76	ALENQHTIDVTDSEMKNLFEVRRQLRENAEDKGNGCFEIFHQCDNNCIE
H4	CH	Ala75	ALENQHTIDVTDSEMDKLFERVRQLRENAEDKGNGCFEIFHQCDNNCIE
H4	RT	NJ85	ALENQHTIDVTDSEMKNLFEVRRQLRENAEDKGNGCFEIFHQCDNNCIE
H4	TU	Min80	ALENQHTIDVTDSEMKNLFEVRRQLRENAEDKGNGCFEIFHQCDNSCIE
H4	SE	Mas82	ALENQHTIDVTDSEMKNLFEVRRQLRENAEDKGNGCFEIFHQCDNKCIE
H1		Khab77	LLENERTLDFHDSNVKNLYEKVKSQKNNAKEIGNGCFEFYHKCNNECME
H1		Kiev5979	LLENERTLDFHDSNVKNLYEKVKSQKNNAKEIGNGCFEFYHKCNNECME
H1		Wsn33	LLENERTLDFHDLNVKNLYEKVKSQKNNAKEIGNGCFEFYHKCDNECME
H1		Pr834	LLENERTLDFHDSNVKNLYEKVKSQKNNAKEIGNGCFEFYHKCDNECME
H1		Swnj1176	LLENERTLDFHDSNVKNLYEKVRSQKNNAKEIGNGCFEFYHKCDDTCME
H1		Hlhmiv	LLENERTLDFHDSNVKNLYEKVKSQKNNAKEIGNGCFEFYHKCDNECME
H2		Jap30757	LMENERTLDFGDSNVKNLYDKVRMQLRDNVKELGNGCFEFYHKCDDCEMN
H2		Jap30557	LMENERTLDFHDSNVKNLYDKVRMQLRDNVKELGNGCFEFYHKCDDCEMN
H5		Chpa83	LMENERTLDLHDSNVKNLYDKVRLQLRDNKEWNGNGCFEFYHKCDNECME
H5		TUIrel137883	LMENERTLDFHDANVKSLYDKVRLQLKDNASELGNGCFEFYHKCDNECME
H5		DUIrel11383	LMENERTLDFHEANVKSLYDKVRLQLKDNASELGNGCFEFYHKCDNECME
H10		Minksw84	AMENQHTIDMADSEMLALYERVRKQLRQNAEEDGKGCFEYHTCDDSCME
H10		Chgern49	AMENQHTIDMADSEMLALYERVRKQLRQNAEEDGKGCFEYHTCDDSCME
H7		Stvic85	AMENQHTIDLTDSEMKNLYERVRQLRENAEEDCTGCFEIFHKCDDDCMA
H7		Chvic85	AMENQHTIDLADSEMKNLYERVRQLRENAEEDCTGCFEIFHKCDDDCMA
H7		Sealma80	AMENQHTIDLADSEMKNLYERVRKQLRENAEEDGTGCFEIFHKCDDQCME
H7		Fpv	AMENQHTIDLADSEMKNLYERVRKQLRENAEEDGTGCFEIFHKCDDDCMA

H3		X31	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Nt68	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Vic375	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Bk179	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Bk279	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Lenin38580	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Eng32177	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Mem10272	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Aichi268	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Nt606829c	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Mem10772	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		Duukr63	SIRNGTYDHDYRDEALNNRFQIKGVELKSGYKDILWISFAISCLLLCV
H3		EQ urg63	SIRNGTYDHDYRDEALNNRFQIRGVELKSGYKDILWISFAISCFLICV
H3		EQ mia63	SIRNGTYDHDYRDEALNNRFQIRGVELKSGYKDILWISFAISCFLICV
H3		EQ tky71	SIRNGTYDHDYRDEALNNRFQIRGVELKSGYKDILWISFAISCFLICV
H3		EQ alg72	SIRNGTYDHDYRDEALNNRFQIRSVELKSGYKDILWISFAISCFLICV
H3		EQ nm 76	SIRNGTYDHYIYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLICV
H3		EQ fon79	SIRNGTYDHYIYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLICV
H3		EQ rom80	SIRNGTYDHYIYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLICV
H3		EQ san85	SIRNGTYDHYIYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLICV
H3		EQ tn 86	SIRNGTYDHYIYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLICV
H3		EQ ky 86	SIRNGTYDHYIYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLICV
H3		EQ ky 87	SIRNGTYDHYIYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLICV
H3		DU hk577	SIRNGTYDHDVYRDESINNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		DU hk880	SIRNGTYDHDYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		DU hk3380	SIRNGTYDHDYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		DU hk782	SIRNGTYDHDYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCA
H3		DU hk2182	SIRNGTYDHDYRDEALNNRFQIKGVELKSGYKNWILWISFAISCFLLCV
H3		DU hk985	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3		DU hk1085	SIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV

H3 SW hk12682	SIRNGTYDHDYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H3 SW hk8178	SIRNGTYDHDYRDEALNNRFQIKGVELKSGYKDILWISFAISCFLLCV
H4 DU Czk56	SIRNGTYDHDYRDEAINNNRFQIQGVKLTQGYKDILWISFSISCFLLLVA
H4 GT Aus79	SIRNGTYDHDYRDEAINNNRFQIQGVKLTQGYKDILWISFSISCFLLLVA
H4 BU Hok77	SIRNGTYDHDYRDEAINNNRFQIQGVKLTQGYKDILWISFSISCFLLLVA
H4 DU N276	SIRNGTYDHDYRDEAINNNRFQIQGVKLTQGYKDILWISFSISCFLLLVA
H4 DU Alb76	SIRNGTYDHDYRDEAINNNRFQIQGVKLTQGYKDILWISFSISCFLLLVA
H4 CH Ala75	SIRNGTYDHDYRDEAINNNRFQIQGVKLTQGYKDILWISFSISCFLLLVA
H4 RT NJ85	SIRNGTYDHDYRDEAINNNRFQIQGVKLTQGYKDILWISFSISCFLLLVA
H4 TU Min80	SIRNGTYDHDYRDEAINNNRFQIQGVKLTQGYKDILWISFSISCFLLLVA
H4 SE Mas82	SIRNGTYDHDYRDEAINNNRFQIQGVKLTQGYKDILWISFSISCFLLLVA
H1 Khab77	SVKNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLCLL
H1 Kiev5979	SVKNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLCLL
H1 Wsn33	SVRNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLL
H1 Pr834	SVRNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLL
H1Swj1176	SVKNGTYDYPKYSEESKLNREEIDGVKLESTRIYQILAIYSTVASSLVLL
H1 Hihmiv	SVRNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLL
H2 Jap30757	SVKNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLL
H2 Jap30557	SVKNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLL
H2 Chpa83	SVRNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLL
H5 TUIre137883	SIRNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLL
H5 DUIre11383	SIRNGTYDYPKYSEESKLNREKIDGVKLESMGVYQILAIYSTVASSLVLL
H10Minkaw84	SIRNNTYDHSQYREEALNRLNINSVKLSSGYKDILWISFSGASCFLLLA
H10Chgern49	SIRNNTYDHSQYREEALNRLNINSVKLSSGYKDILWISFSGASCFLLLA
H7 Stvic85	SIRNNTYDHSQYREEAMQNRVKIDPVKLSSGYKDILWISFSGASCFLLLA
H7 Chvic85	SIRNNTYDHSQYREEAMQNRVKIDPVKLSSGYKDILWISFSGASCFLLLA
H7Sealma80	SIRNNTYDHTQYRAKSLQNRQIDPVKLSSGYKDILWISFSGASCFLLLA
H7 Fpv	SIRNNTYDHSKYREEAMQNRQIDPVKLSSGYKDILWISFSGASCFLLLA

H3 X31	VLL-GFIMWACQKGNIRCNICI
H3 Nt68	VLL-GFIMWACQKGNIRCNICI
H3 Vic375	VLL-GFIMWACQKGNIRCNICI
H3 Bk179	VLL-GFIMWACQKGNIRCNICI
H3 Bk279	VLL-GFIMWACQKGNIRCNICI
H3Lenin38580	VLL-GFIMWACQKGNIRCNICI
H3 Eng32177	VLL-GFIMWACQKGNIRCNICI
H3 Mem10272	VLL-GFIMWACQKGNIRCNICI
H3 Aich1268	VLL-GFIMWACQKGNIRCNICI
H3 Nt606829c	VLL-GFIMWACQKGNIRCNICI
H3 Mem10772	VLL-GFIMWACQKGNIRCNICI
H3 Duukr63	VLL-GFIMWACQKGNIRCNICI
H3 EQ urg63	VLL-GFIMWACQKGNIRCNICI
H3 EQ mia63	VLL-GFIMWACQKGNIRCNICI
H3 EQ tky71	VLL-GFIMWACQKGNIRCNICI
H3 EQ alg72	VLL-GFIMWACQKGNIRCNICI
H3 EQ nm 76	VLL-GFIMWACQKGNIRCNICI
H3 EQ fon79	VLL-GFIMWACQKGNIRCNICI
H3 EQ rom80	VLL-GFIMWACQKGNIRCNICI
H3 EQ san85	VLL-GFIMWACQKGNIRCNICI
H3 EQ tn 86	VLL-GFIMWACQKGNIRCNICI
H3 EQ ky 86	VLL-GFIMWACQKGNIRCNICI
H3 EQ ky 87	VLL-GFIMWACQKGNIRCNICI
H3 DU hk577	VLL-GFIMWACQKGNIRCNICI
H3 DU hk880	VLL-GFIMWACQKGNIRCNICI
H3 DU hk3380	VLL-GFIMWACQKGNIRCNICI
H3 DU hk782	VLL-GFIMWACQKGNIRCNICI
H3 DU hk2182	VLL-GFIMWACQKGNIRCNICI
H3 DU hk985	VLL-GFIMWACQKGNIRCNICI

H3 DU hk1085	VLL-GFIMWACQRGBNIRCNICI
H3 SW hk12682	VLL-GFIMWACQRGBNIRCNICI
H3 SW hk8178	VLL-GFIMWACQRGBNIRCNICI
H4 DU Czk56	LLL-AFILWACQNGNIRCQICI
H4 GT Aus79	LLL-AFILWACQNGNIRCQICI
H4 BU Hok77	LLL-AFILWACQNGNIRCQICI
H4 DU NZ76	LLL-AFILWACQNGNIRCQICI
H4 DU Alb76	LLL-AFILWACQNGNIRCQICI
H4 CH Ala75	LLL-AFILWACQNGNIRCQICI
H4 RT NJ85	LLL-AFILWACQNGNIRCQICI
H4 TU Min80	LLL-AFILWACQNGNIRCQICI
H4 SE Mas82	LLL-AFILWACQKGNIRCQICI
H1 Khab77	VSLGAISFWMCNNGSLQCRICI
H1 Kiev5979	VSLGAISFWMCNNGSLQCRICI
H1 Wsn33	VSLGAISFWMCNNGSLQCRICI
H1 Pr834	VSLGAISFWMCNNGSLQCRICI
H1Swj1176	VSLGAISFWMCNNGSLQCRICI
H1 Hlhmiv	VSLGAISFWMCNNGSLQCRICI
H2 Jap30757	IMMAGISFWMCNNGSLQCRICI
H2 Jap30557	IMMAGISFWMCNNGSLQCRICI
H5 Chpa83	IMVAGLSFWMCNNGSLQCRICI
H5 TUIrel137883	IMIAGLSFWMCNNGSLQCRICI
H5 DUIrel1383	IMIAGLSFWMCNNGSLQCRICI
H10Minksw84	AVMGLVFF-CLKNGNMQCTICI
H10Chgern49	VVMGLVFF-CLKNGNMRCCTICI
H7 Stvic85	IAMGLVF-MCVKNGNMRCCTICI
H7 Chvic85	IAMGLVF-MCVKNGNMRCCTICI
H7Sealma80	IAMGLVF-ICIKNGNMRCCTICI
H7 Fpv	IAVGLVF-ICVKNGNMRCCTICI

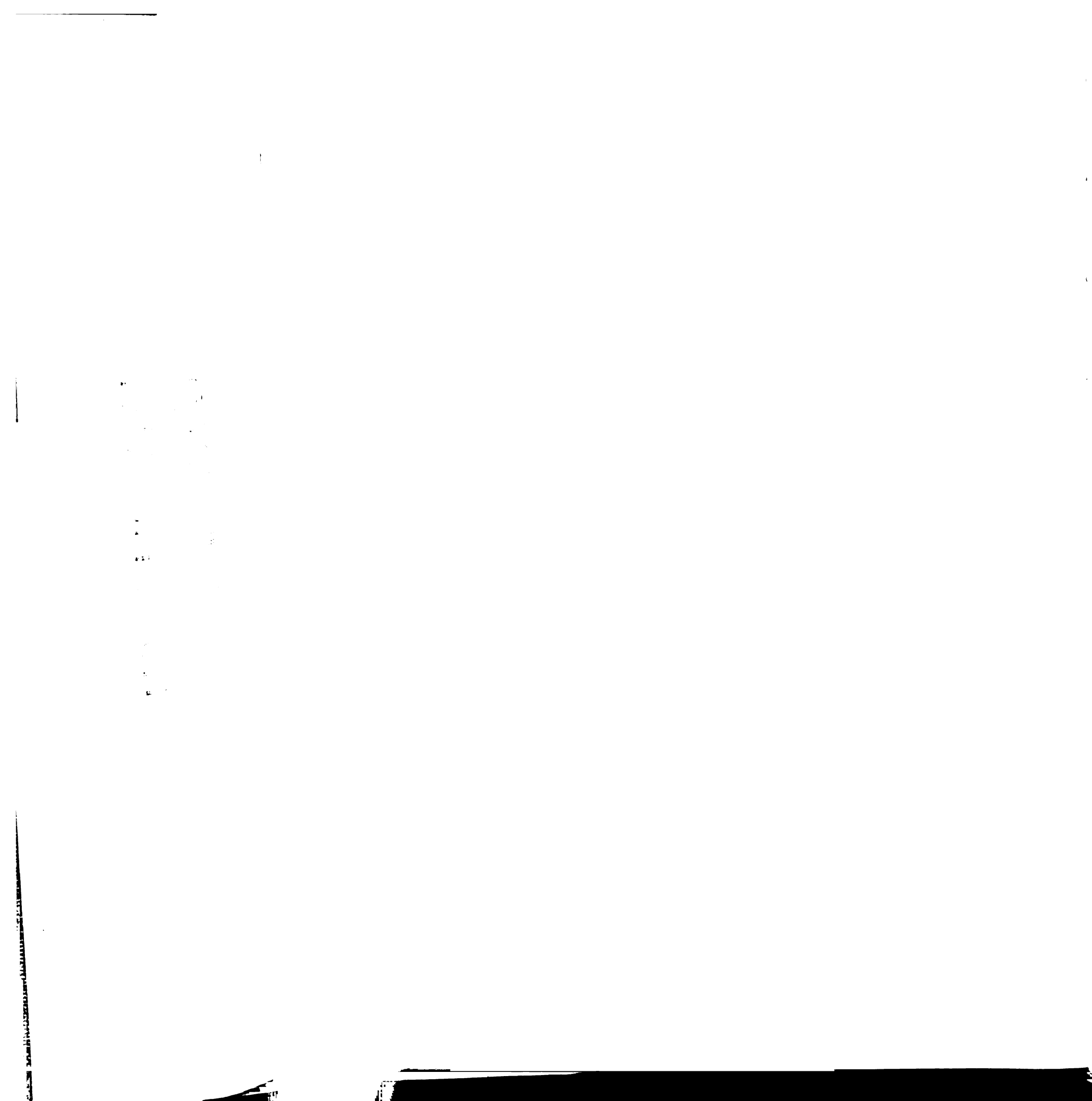
References

1. Kawaoka, Y., Naeve, C. W. and Webster, R. G., "Is virulence of H5N2 influenza viruses in chickens associated with loss of carbohydrate from the hemagglutinin?" *Virology* 139 303-316 (1984).
2. Devereux, J., Haeberli, P. and Smithies, O., "A comprehensive set of sequence analysis programs for the VAX." *Nucl Acids Res* 12 387-395 (1984).

3. Gething, M.-J., Bye, J., Skehel, J. and Waterfield, M., "Cloning and DNA sequence of double-stranded copies of hemagglutinin genes from H2 and H3 strains elucidates antigenic shift and drift in human influenza virus." *Nature* 287 301-306 (1980).
4. Wilson, I. A., personal communication.
5. Jou, W. M., Verhoeven, M., Devos, R., Saman, E., Fang, R., Huylebroeck, D., Fiers, W., Threlfall, G., et al., "Complete structure of the hemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA." *Cell* 19 683-696 (1980).
6. Both, G. W. and Sleight, M. J., "Conservation and variation in the hemagglutinins of Hong Kong subtype influenza viruses during antigenic drift." *J Virol* 39 663-672 (1981).
7. Migunova, B. V., Plusnin, A. Z., Petrov, N. A., Rivkin, M. I., Grinbaum, E. B., Obuchova, L. V. and Kuznetzov, O. K., "Primary structure of the gene coding for the hemagglutinin of influenza virus A/Leningrad/385/80(H3N2): detection of a point mutation responsible for the antigenic drift." *Acta virol* 24 209-218 (1990).
8. Hauptmann, R., Clarke, L. D., Mountford, R. C., Bachmayer, H. and Almond, J. W., "Nucleotide sequence of the hemagglutinin gene of influenza virus A/England/321/77." *J Gen Virol* 64 215-220 (1983).
9. Laver, G. and Air, G. (Eds.), Structure and Variation in Influenza Virus. (Elsevier/North-Holland, New York, 1980).
10. Palese, P. and Kingsbury, D. W. (Eds.), Genetics of Influenza Viruses. (Springer-Verlag, New York, 1983).

11. Kawaoka, Y., Bean, W. J. and Webster, R. G., "Evolution of the hemagglutinin of equine H3 influenza viruses." *Virology* 169 283-292 (1989).
12. Kida, H., Kawaoka, Y., Naeve, C. W. and Webster, R. G., "Antigenic and genetic conservation of H3 influenza virus in wild ducks." *Virology* 159 109-119 (1987).
13. Kida, H., Shortridge, K. F. and Webster, R. G., "Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China." *Virology* 162 160-166 (1988).
14. Donis, R. O., Bean, W. J., Kawaoka, Y. and Webster, R. G., "Distinct lineages of influenza virus H4 hemagglutinin genes in different regions of the world." *Virology* 169 408-417 (1989).
15. Beklemishev, A. B., Blinov, V. M., Vasilenko, S. K., Golovin, S. Y., Karginov, V. A., Mamaev, L. V., Netesov, S. V., Petrov, N. A., et al., "Primary structure of a full-length DNA copy of the hemagglutinin gene of the influenza virus A/Kiev/59/79 (H1N1)." *Bioorg Khim* 12 375-381 (1986).
16. Winter, G., Fields, S. and Brownlee, G. G., "Nucleotide sequence of the hemagglutinin gene of a human influenza virus H1 subtype." *Nature* 292 72-75 (1981).
17. Caton, A. J., Brownlee, G. G., Yewdell, J. W. and Gerhard, W., "The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype)." *Cell* 31 417-427 (1982).
18. Robertson, J. S., "Sequence analysis of the hemagglutinin of A/Taiwan/1/86, a new variant of human influenza A (H1N1) virus." *J Gen Virol* 68 1205-1208 (1987).

19. Kawaoka, Y., Nestorowics, A., Alexander, D. J. and Webster, R. G., "Molecular analyses of the hemagglutinin genes of H5 influenza viruses: Origin of a virulent turkey strain." *Virology* 158 218-227 (1987).
20. Feldmann, H., Kretzschmar, E., Klingeborn, B., Rott, R., Klenk, H.-D. and Garten, W., "The structure of serotype H10 hemagglutinin of influenza A virus: comparison of an apathogenic avian and a mammalian strain pathogenic for mink." *Virology* 165 428-437 (1988).
21. Nestorowicz, A., Kawaoka, Y., Bean, W. J. and Webster, R. G., "Molecular analysis of the hemagglutinin genes of Australian H7N7 influenza viruses: Role of passerine birds in maintenance or transmission?" *Virology* 160 411-418 (1987).
22. Naeve, C. W. and Webster, R. G., "Sequence of the hemagglutinin gene from influenza virus A/Seal/Mass/1/80." *Virology* 129 298-308 (1983).
23. Air, G. M., "Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A virus." *PNAS* 78 7639-7643 (1981).



Appendix C

Program Listings

DIFFDOCK (Chapter 10) is a modification of the DOCK2.0 program [1]. Routines specific for DIFFDOCK are listed below in their entirety. **DIFFDOCK.BLK** contains variable declarations, **clurd.f** reads the receptor spheres cluster with flagged spheres, and **difchk.f** counts the number of labeled spheres in each match. Subroutines **mainnew.dfdk.f**, **single.dfdk.f**, and **search.dfdk.f** are slight modifications of the corresponding DOCK2.0 routines **mainnew.f**, which reads input, **single.fan.f**, which docks one ligand, and **search.fan.f**, which searches a database of small molecules. Differences between the original routines and the **diffdock** routines recognized by the UNIX utility "diff" are listed below. Lines preceded by '<' are specific to DIFFDOCK. Surrounding lines and explanatory remarks are included for context. The complete subroutine **ilgen.dfdk.f**, which selects ligand and receptor bins (sets of atoms or spheres) for matching, is also listed. Following the DIFFDOCK subroutines is the code for **ezdock3.f**, an interactive program that creates the INDOCK input file for DOCK3.0 [2].

100

1

DIFDOCK.BLK

```
      INTEGER MAXSET
      PARAMETER (MAXSET = 10)
      INTEGER DMAX
      PARAMETER (DMAX = 200)
c     dmax should be the same as maxpts

      CHARACTER*1 DIFTMP
      CHARACTER*14 ADUM14
      INTEGER DIFSET, NSET, NSETREQ
      INTEGER NDIF(MAXSET), NDIFREQ(MAXSET)
      INTEGER DIFFSPHERES(MAXSET), NTOT
      INTEGER DIFMODE, DIFSPH(DMAX)

      COMMON /DIFDAT1/ DIFTMP
      COMMON /DIFDAT2/ DIFSET, NDIF, NSET, NDIFREQ,
& DIFFSPHERES, NTOT, DIFMODE, DIFSPH, NSETREQ

c     DIFSPH(i): set number for a diffdock sphere, 0 else
c     ndif: number of spheres labeled for differential docking
c     for set maxset
c     number of diffdock sets for each set maxset required to be
c     used in a match.
c     nset is number of sets used.
c     for now use highest number of difset
c     difmode selects whether screening for a diffdock sphere will
c     be done before or after score and orient.
```

clurd.f

```
c READ IN RECEPTOR SPHERE CLUSTER
c if there is a 'D' after the coords, this
c sphere is flagged to be used in differential docking.
c if there is a 'D' and no number (or too high), assume set 1
c if no D, then assign set 0 (no diffdock set)

      DO 11 I = 1, MAXSET
          NDIF(I) = 0
          NDIFREQ(I) = 0
11     CONTINUE
      NSET = 0
      NSETREQ = 0

      DO 60 I = 1, NSPH
          DIFSPH(I) = 0
          READ(1, 58) ICLUS(I), (SPCORR(J, I), J = 1, 3), ADUM14, DIFTMP, DIFSET
          IF ( DIFMODE .GT. 0 .AND. DIFTMP .EQ. 'D') THEN
              IF (DIFSET .LT. 1 .OR. DIFSET .GT. MAXSET) DIFSET = 1
              DIFSPH(I) = DIFSET
              NDIF(DIFSET) = NDIF(DIFSET) + 1
              WRITE(6, '(A,1x,I3,1x,A,1x,I3,1x,A,1x,I5)') 'Sphere',
& ICLUS(I), 'Diffdock set ', DIFSET, 'number ', NDIF(DIFSET)
```

1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 26

```

        IF (DIFSET .GT. NSET) NSET = DIFSET
    ENDIF
60    CONTINUE
58    FORMAT(i5,3f10.5,A14,A1,I3)

    IF (DIFMODE .GT. 0) THEN
        READ(51,*) NSETREQ
        WRITE(6,*) 'NUMBER OF SETS REQUIRED: ',NSETREQ
        WRITE(6,*) 'NUMBER OF DIFFDOCK SPHERE SETS READ: ',NSET
        if (nsetreq .gt. nset) then
            write(6,*) 'More sets required than present in file.'
            stop
        endif
        DO 111 i = 1,NSET
            READ(51,*,end=112) ii,j
            NDIFREQ(ii) = j
            WRITE(6,*) 'NUMBER OF REQUIRED SPHERES FROM SET ', ii,
&                ' : ',NDIFREQ(ii)
111        CONTINUE
112        CONTINUE
    ENDIF

c    Make sure there are at least as many spheres marked in a
c    given set as are required in that set.

    ntot = 0
    do 78 i = 1, nset
        write(6,*) 'nset',i,' ndif',ndif(i),' reqs',ndifreq(i)
        ntot = ntot + ndifreq(i)
        if (ndif(i) .lt. ndifreq(i) ) then
            write(6,*) 'Set ',i,' requires more spheres in match'
            write(6,*) 'than are labeled. ',ndif(i),ndifreq(i)
            write(6,*) 'program stops'
            stop
        endif
78    continue
    if (ntot .gt. nodlim) then
        write(6,*) 'more spheres required then nodlim',ntot,nodlim
        write(6,*) 'program stops'
        stop
    endif

```

difchk.f

c Check that the match contains the correct number of diffdock spheres.
c Diffdockv1, with sets
c DLB July 25 1991
c Each diffdock sphere can only be in one set.

c difsph(i) gives the set number that diffdock sphere i is in (else 0).

```

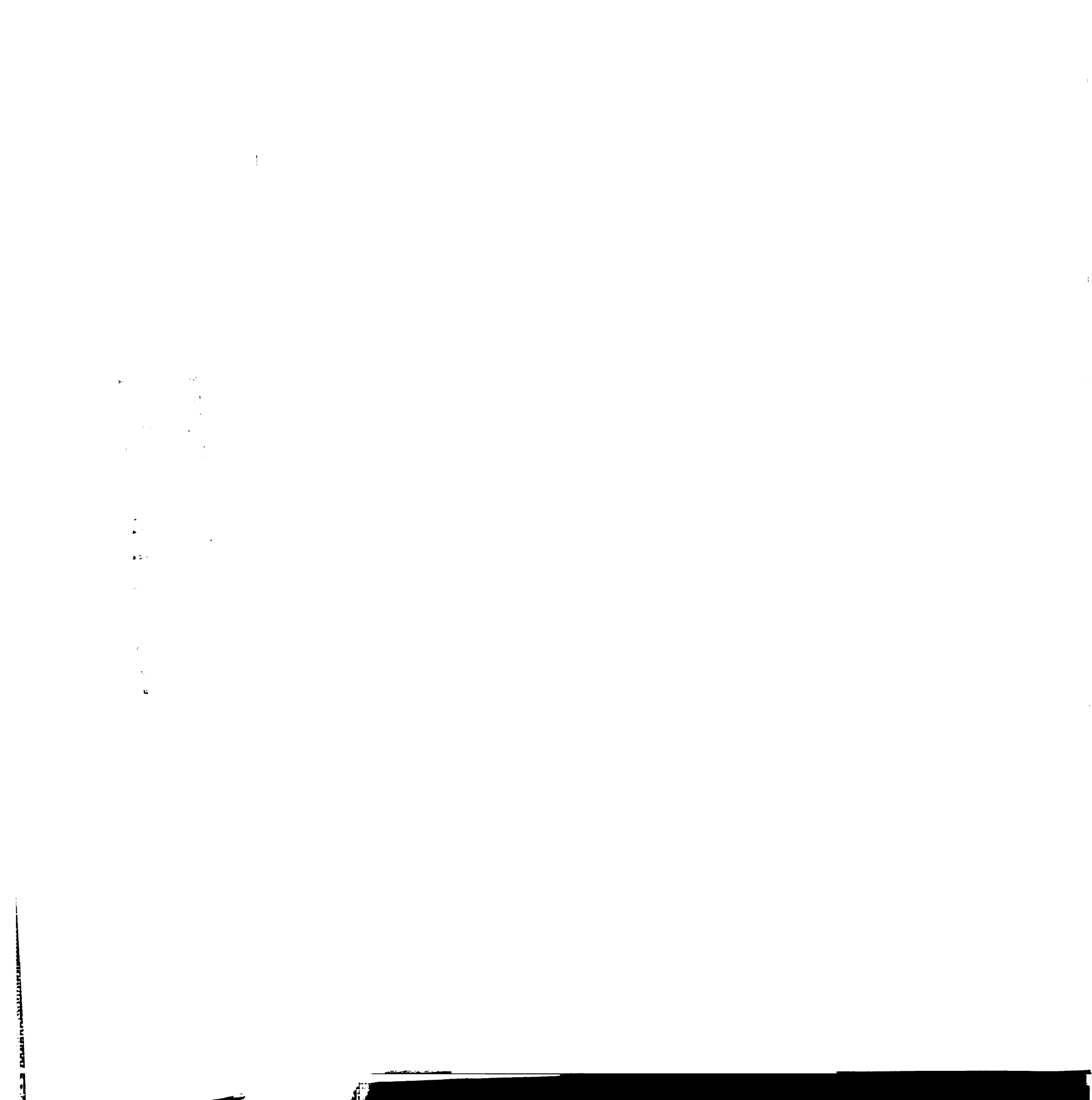
subroutine difchk(wrtnod,iasign,iwrite,MAXPTS)

```

```

integer id,wrtnod

```



```

integer iassign(100,2)
logical iwrite
integer MAXPTS,setno
INCLUDE 'DIFFDOCK.BLK'

if (difmode .ne. 1) return

do 10 id = 1,nset
  diffspheres(id) = 0
10 continue

do 20 id = 1,wrtnod+1
  setno = difsph(iassign(id,2))
  if (setno .gt. 0) then
    diffspheres(setno) = diffspheres(setno) + 1
  endif
20 continue

do 30 id = 1,nset
  if (diffspheres(id) .lt. ndifreq(id)) then
    iwrite = .FALSE.
    return
  endif
30 continue

return
end

```

mainnew.dfdk.f

Variable Declarations.

57,58d55

```

<
<      INCLUDE 'DIFFDOCK.BLK'

```

Determination of run mode: differential or not.

308,309c310

```

<      READ(51,*)DIFMODE
<      IF (DIFMODE .LT. 1 .OR. DIFMODE .GT. 2) DIFMODE = 0

```

523c524,527

```

<      IF(DIFMODE.GT.0) THEN

```

525,526c529,530

```

<      WRITE(6,'(/,T10,A)') 'THIS IS A DIFFDOCK RUN'
<      WRITE(6,*) 'DIFFDOCK MODE SELECTED: ',DIFMODE

```

Read receptor sphere cluster.

620,623c624,634

```

<
<      INCLUDE 'CLURD.F'

```


<
628c639

< read(1,59) idum,dum,dum,dum

630d640

< 59 format(i5,3f10.5)

Modified calls to single and search routines.

655c665

666c676

```
      IF(MODE.EQ.'SINGLE') THEN
        CALL SINGLE(NSPH,SPCORR,DISR,DISL,DMAXR,NATR,CR,DISLIM,
& NODLIM,LOWNOD,CONCUT,DMIN,DISCUT,RATIOM,NSAV,LIGTYP,
& ICTBMP,LREAD,MAXDIX,MAXDIY,MAXDIZ,GRID,MAXPTS,MAXWID,
& MAXNOD,LIGNUM,SPHNUM,RTOTBN,LTOTBN,SPTMP,LGTEMP,SBINSZ,
& LBINSZ,SOVLAP,LOVLAP,LIGTMP,SPHTMP,DISTMP,SDTMP,PROF,
< & NODUSD,MINSR,USRDEV,LRMSD,TMPBIN,LIGEXP,
& SPHEXP,FRATIO,EXPMAX,FCTBMP)
      ENDIF

      IF(MODE.EQ.'SEARCH') THEN
        CALL SEARCH(NSPH,SPCORR,DISR,DISL,DMAXR,NATR,CR,DISLIM,
& NODLIM,LOWNOD,CONCUT,DMIN,DISCUT,NATMIN,NATMAX,RATIOM,NSAV,
& NNSAV,IRESR,MOLTOT,MOLSAV,INCHYD,ICTBMP,LREAD,MAXDIX,MAXDIY,
& MAXDIZ,GRID,MAXPTS,MAXWID,MAXNOD,LIGNUM,SPHNUM,RTOTBN,LTOTBN,
& SPTMP,LGTEMP,SBINSZ,LBINSZ,SOVLAP,LOVLAP,LIGTMP,SPHTMP,
< & DISTMP,SDTMP,PROF,NODUSD,TMPBIN,LIGEXP,SPHEXP,
& FRATIO,EXPMAX,FCTBMP)
      ENDIF
```

single.dfdk.f

Modified arguments.

6,7c6,7

```
      SUBROUTINE SINGLE(NSPHR,SPCORR,DISR,DISL,DMAXR,NATR,CR,DISLIM
& ,NODLIM,SLOWND,CONCUT,DMIN,DISCUT,RATIOM,NSAV,LIGTYP,ICTBMP,LREAD
& ,MAXDIX,MAXDIY,MAXDIZ,GRID,MAXPTS,MAXWID,MAXNOD,LIGNUM,SPHNUM
& ,RTOTBN,LTOTBN,SPTMP,LGTEMP,SBINSZ,LBINSZ,SOVLAP,LOVLAP,LIGTMP
< & ,SPHTMP,DISTMP,SDTMP,PROF,NODUSD,MINSR,USRDEV,LRMSD
< & ,TMPBIN,LIGEXP,SPHEXP,FRATIO,EXPMAX,FCTBMP)
```

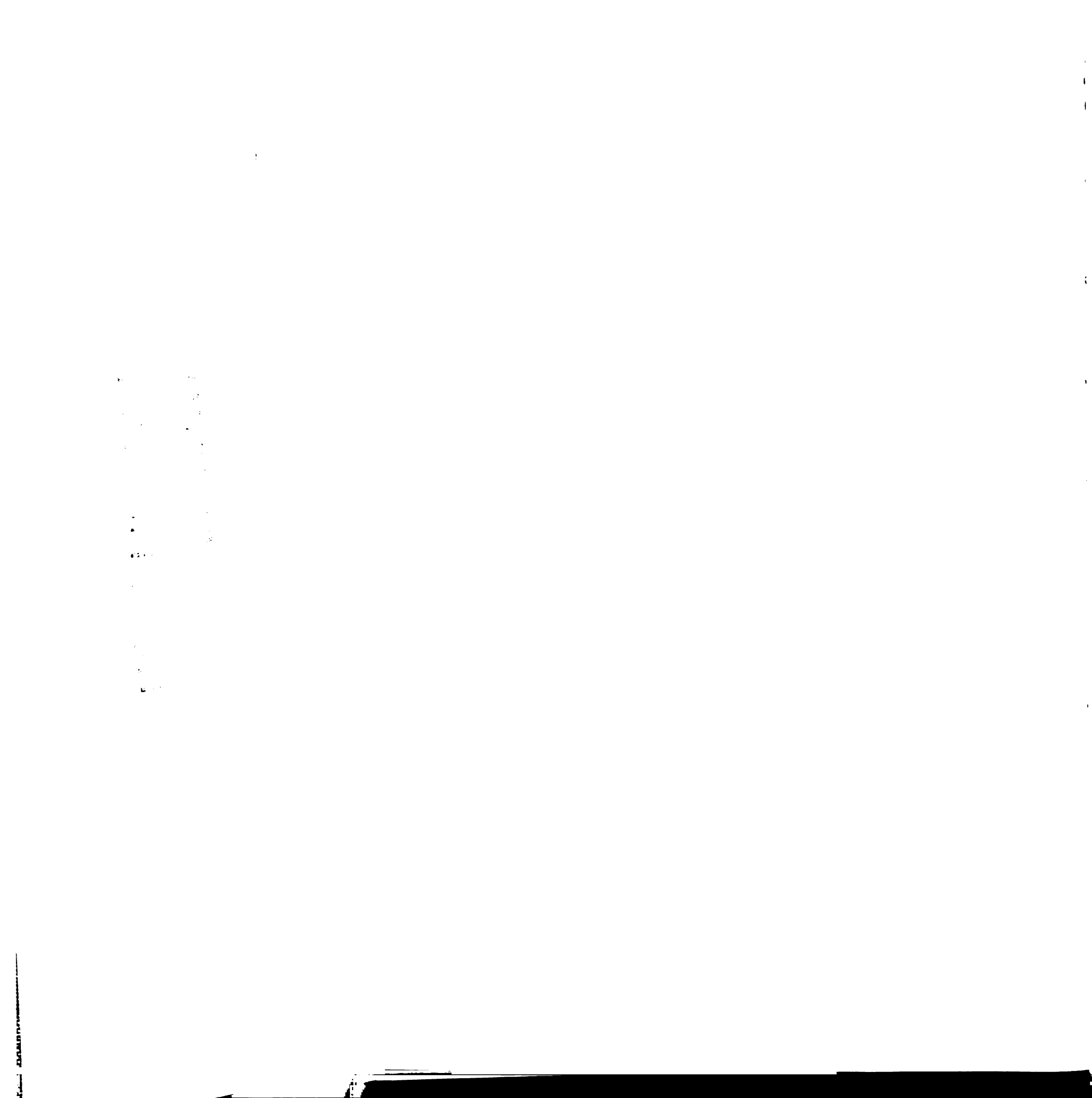
This line deleted to reflect change to global variables.

286a287,289

> LOGICAL DIFSPH(MAXPTS), DIFDCK

Modified call to profil.

460,461c463,464



```

        IF (PROF) THEN
            CALL PROFIL (NSPHR, SPCORR, DISR, DISL, DMAXR, NATR, CR, DISLIM
&            , NODLIM, SLOWND, CONCUR, DMIN, DISCUT, RATIO, NSAV, LIGTYP
&            , ICTBMP, LREAD, MAXDIX, MAXDIY, MAXDIZ, GRID, MAXPTS, MAXWID
&            , MAXNOD, LIGNUM, SPHNUM, RTOTBN, LTOTBN, SPTEMP, LGTEMP, SBINSZ
&            , LBINSZ, SOVLAP, LOVLAP, LIGTMP, SPHTMP, DISTMP, SDTMP
<        &            , PROF, NODUSD, MINSCR, USRDEV, LRMSD
<        &            , TMPBIN, LIGEXP, SPHEXP, FRATIO, EXPMAX
&            , CS, CL, NATL, NSPHL)
            RETURN
        ENDIF

```

Modified call to ilgen:

484c487

```

        CALL ILGEN1 (CENTRL, NODLIM, CENTRR, RTOTBN, LTOTBN, LIGNUM,
&        SPHNUM, ILIST, LGTEMP, SPTEMP, ILMAX, NLIST, MAXPTS,
<        &        MAXNOD, MAXWID, DIVLIG, DIVSPH, LOWNOD)

```

Test match for use of the required differential spheres.

569, 573d571

```

<
<        IF (WRTNOD.GE.NODLIM.AND.IWRITE) THEN
<        CALL DIFCHK (WRTNOD, IASIGN, IWRITE, MAXPTS)
<        ENDIF
<
636, 639d633

```

```

<        IF (WRTNOD.GE.lownod.AND.jWRITE.and.wrtnod.ge.jcnt-1) THEN
<        CALL DIFCHK (WRTNOD, IASIGN, jWRITE, MAXPTS)
<        ENDIF
<

```

search.dfdk.f

Modified arguments.

7c7, 8

```

SUBROUTINE SEARCH (NSPHR, SPCORR, DISR, DISL, DMAXR, NATR, CR, DISLIM,
&NODLIM, SLOWND, CONCUR, DMIN, DISCUT, NATMIN, NATMAX, RATIO, NSAV, NNSAV,
&IRESTR, MOLTOT, MOLSAV, INCHYD, ICTBMP, LREAD, MAXDIX, MAXDIY, MAXDIZ,
&GRID, MAXPTS, MAXWID, MAXNOD, LIGNUM, SPHNUM, RTOTBN, LTOTBN, SPTEMP,
&LGTEMP, SBINSZ, LBINSZ, SOVLAP, LOVLAP, LGNTMP, SPHTMP, DISTMP, SDTMP,
< &PROF, NODUSD, TMPBIN, LIGEXP, SPHEXP, FRATIO, EXPMAX, FCTBMP)

```

Delete this line from the DOCK2 program to reflect change to global variables in DIFFDOCK.

318a320, 322

```

>        LOGICAL DIFSPH (MAXPTS), DIFDCK

```

DOCK2 bug fix.

538,541d540

```
<c   time for a new set of ligand bins.
<   IF (MOD (IMAIN, NSPHR).EQ.0) CENTRL=CENTRL-1
```

Modified call to ilgen.

544c543

```
      CALL ILGEN1 (CENTRL, NODLIM, CENTRR, RTOTBN, LTOTBN, LIGNUM,
&SPHNUM, ILIST, LGTEMP, SPTEMP, ILMAX, NLIST, MAXPTS,
<    &MAXNOD, MAXWID, DIVLIG, DIVSPH, LOWNOD)
```

Before writing a match test for required differential spheres.

628,632d626

```
<   IF (WRTNOD.GE.NODLIM.AND.IWRITE) THEN
<     CALL DIFCHK (WRTNOD, IASIGN, IWRITE, MAXPTS)
<   ENDIF
```

675,678d668

```
<   IF (WRTNOD.GE.lownod.AND.jWRITE.and.wrtnod.ge.jcnt-1) THEN
<     CALL DIFCHK (WRTNOD, IASIGN, jWRITE, MAXPTS)
<   ENDIF
```

ilgen.dfdk.f

```
*****
*-----ILGEN-----*
```

cdb Start with first ligand bin. Search receptor bins from longest
cdb distance to shortest until find one with matching distance label.
cdb Save the lig bin - rec bin pair in ilist, then go to the next ligand
cdb bin. Start looking for its corresponding bin in the receptor with
cdb the bin after the one that was just used. If no matching bin is
cdb found then Note: ilist array is started with index 2 to be
cdb consistent with rest of dock v2.0. When that's fixed, fix it here
cdb too. Note: right now ilist includes as many ligbin - rec bin
cdb partners as possible. May need to change dimensions of ilist
cdb (lismax). If fewer than lownod pairs are found, ilist is set to all
cdb zeroes. Includes bounds checking to make sure there aren't too many
cdb pairs.

cdb May 31 1990 to make the routine consistent with the rest of the
cdb currentcode set ilist to all 0's if fewer than nodlim pairs are
cdb found.

CCC Original code of Brian Shoichet extensively modified by Dale Bodian.

```
      SUBROUTINE ILGEN1 (CENTRL, NODLIM, CENTRR, RTOTBN, LTOTBN, LIGNUM,
&    SPHNUM, ILIST, LGTEMP, SPTEMP, ILMAX, NLIST, MAXPTS,
&    MAXNOD, MAXWID, DIVLIG, DIVSPH, LOWNOD)
```

```
      INTEGER MAXPTS, MAXNOD, MAXWID
CCC   array bounds, set as parameters above.
      INTEGER I, J, K, KK, L
```



```

CCC    loop variables all.
        INTEGER CENTRL,CENTRR
CCC    centrr: which receptor center bins are made from.
CCC    centrl: which ligand center bins are made from
        INTEGER NODLIM,ILMAX,NLIST
CCC    nodlim: max. nodes in a sub-graph.
CCC    ilmax is the total number of ligand/sphere node pairs in
CCC    ILIST.  nlist is the maximum ilmax.
        INTEGER LOWNOD
CCC    lownod: min. allowed nodes in a sub-graph.
        REAL DIVLIG,DIVSPH
CCC    divlig,divsph: variable used to match bin distance labels.
        INTEGER RTOTBN(MAXPTS),LTOTBN(MAXPTS),LSTOP,SSTOP
CCC    rtotbn holds the number of bins for any given receptor center.
CCC    ltotbn holds the number of bins for any given ligand center.
CCC    sstop is a temp for a particular center value of rtotbn.
CCC    lstop is a temp for a particular center value of ltotbn.
        INTEGER LIGNUM(MAXPTS,0:MAXNOD,MAXWID)
        INTEGER SPHNUM(MAXPTS,0:MAXNOD,MAXWID)
CCC    lignum and sphnum hold the nodes of each ligand and receptor
CCC    center, respectively.  The nodes are organized into bins, as per
CCC    DISCRETE.  The first dimension holds the original center, the
CCC    second column holds the bins and the third column holds the
CCC    nodes in the particular bins.  note that the first and second
CCC    positions in the third dimension contain the bin label and
CCC    how many nodes it contains.
        INTEGER ILIST(30,2)
CCC    pointer to the bins to draw the nodes from for matching.
        INTEGER SPTEMP(MAXNOD),LGTEMP(MAXNOD), NODCNT
CCC    temporary arrays used to store nodes as ilist is being built.
CCC    nodcnt: temporary counter variable used as ilist is built.
CCC    indicates size of sub-clique.
        integer lbin,rbin
cdb    lbin is the ligand bin for which a corresponding receptor bin
cdb    is needed.
cdb    rbin is the receptor bin that's currently being considered as
cdb    a partner for lbin.
        integer lismax
cdb    lismax is the maximum size of ilist

cdb initialization

        lismax = 30
        nodcnt = 0
        lbin = 0
        rbin = 1
        do 100 k = 1,lismax
            ilist(k,1) = 0
            ilist(k,2) = 0
100    continue

        do 500 while ( lbin .lt. ltotbn(centrl) )
            lbin = lbin + 1
            do 250 while ( int(divlig*real(lignum(centrl,lbin,1))) .lt.
& int(divsph*real(sphnum(centrr,rbin,1))) )
                if ( rbin .eq. rtotbn(centrr) ) then
                    goto 1000
                endif
            enddo
        enddo

```

1. The first part of the document is a list of the names of the persons who were present at the meeting. The names are listed in alphabetical order.

2. The second part of the document is a list of the topics that were discussed at the meeting. The topics are listed in alphabetical order.

```

        rbin = rbin + 1
250  continue
        if ( int(divlig*real(lignum(centrl,lbin,1))) .eq.
& int(divsph*real(sphnum(centrr,rbin,1))) ) then

CCC      use bins w/maximum numbers of centers in a resolution range

        CALL MAXBIN(LIGNUM,LBIN,DIVLIG,MAXNOD,MAXPTS,CENTRL)
        CALL MAXBIN(SPHNUM,RBIN,DIVSPH,MAXNOD,MAXPTS,CENTRR)
        nodcnt = nodcnt + 1
        ilist(nodcnt+1,1) = lbin
        ilist(nodcnt+1,2) = rbin
        rbin = rbin + 1

cdb      bounds check: if start ilist from index 1 (instead of 2)
cdb      then change lismax to lismax - 1.

        if (nodcnt .ge. lismax-1 ) goto 1000
        endif
        500 continue
        1000 continue

cdb check for lownod (or nodlim)
cdb to make it consistent with current version of the bin code (may 90)
cdb only have nonzero ilist if there are at least nodlim pairs.

        IF(NODCNT.Le.NODLIM)THEN
CCC      if not enough bins w/correct distance labels, set
CCC      ilist pointers to empty bin - results in no matches.
        DO 80 K=1,NODLIM
            ILIST(K+1,1)=0
            ILIST(K+1,2)=0
        80  CONTINUE
        ENDIF

        RETURN
        END

```


ezdock3.f

c program ezdock3.f
c April 2 1992 dlb
c Runs interactively to create the INDOCK file required for a
c DOCK3.0 run.

```
character*1 ligtyp,qrmsd,dum1,cont
character*6 mode
character*80 charin,phifil,vdwfil,prefix
character*80 clufil,cluline,mapnam
character*80 ligfil,outfil,lgsfil,fixfil
integer intl,nodlim,ictbmp,inchyd,expmax,lownod
integer natmin,natmax,nsav,nnsav,irestr,moltot,molsav
integer molskip,intrp
integer charlen,fctbmp,versn
real dislim,ratiom, realdum,vdwmax
real lbinsz,lovlap,sbinsz,sovlap
real minscr,usrdev,maxscr,vscale,escale
logical exs

2  FORMAT(A1)
3  FORMAT(A6)
5  FORMAT(I4)
6  FORMAT(4I4)
7  FORMAT(3F10.3)
8  FORMAT(4F10.3)
9  FORMAT(A80)

WRITE(6,*) 'WELCOME TO EZDOCK3,'
WRITE(6,*)
WRITE(6,*) 'AN INTERACTIVE PROGRAM THAT CREATES THE INPUT FILE'
WRITE(6,*) 'INDOCK REQUIRED TO RUN DOCK3.0'
WRITE(6,*)
WRITE(6,*) 'TYPICAL VALUES ARE PROVIDED FOR MOST VARIABLES AS'
WRITE(6,*) 'DEFAULT VALUES.'
WRITE(6,*) 'FOR MORE HELP, SEE THE DOCUMENTATION UNDER THE'
WRITE(6,*) 'VARIABLE NAME GIVEN AT THE PROMPT.'
WRITE(6,*)
WRITE(6,*)

WRITE(6,*) 'PRIOR TO RUNNING DOCK YOU NEED TO HAVE: '
WRITE(6,*)
WRITE(6,*) 'THE SPHERE CLUSTER FILE FROM SPHGEN '
WRITE(6,*) '          and'
WRITE(6,*) 'THE APPLICABLE SCORING GRID(S): '
WRITE(6,*) '          A. CONTACT SCORING GRID FROM DISTMAP '
WRITE(6,*) '          B. FORCE FIELD SCORING GRID FROM CHEMGRID '
WRITE(6,*) '          C. DELPHI SCORING GRID '
WRITE(6,*)
WRITE(6,*)
WRITE(6,*(A,$)) ' CONTINUE? [y] '
READ(5,*(A)) cont
IF (cont .ne. 'y' .and. cont .ne. 'Y' .and. cont .ne. ' ') THEN
    WRITE(6,*) 'PROGRAM STOPS'
    STOP
ENDIF
```

```

INQUIRE(FILE='INDOCK',EXIST=EXS)
IF(EXS .EQ. .TRUE.) THEN
    WRITE(6,*)
    WRITE(6,*) 'FILE INDOCK EXISTS.  OVERWRITE? [y] or n'
    READ(5,2) DUM1
    IF (DUM1 .EQ. 'n' .OR. DUM1 .EQ. 'N') THEN
        WRITE(6,*) 'PROGRAM STOPS.'
        STOP
    ELSE
        OPEN(UNIT=1,FILE='INDOCK',STATUS='OLD')
        REWIND(1)
    ENDIF
ELSE
    OPEN(UNIT=1,FILE='INDOCK',STATUS='NEW')
ENDIF

WRITE(6,*)
WRITE(6,*) 'Select scoring option: '
WRITE(6,*) '      1. CONTACT SCORING ONLY'
WRITE(6, '(A,A)') '      2. CONTACT SCORING PLUS DELPHI ',
& 'ELECTROSTATIC SCORING'
WRITE(6, '(A,A)') '      3. CONTACT SCORING PLUS FORCE ',
& 'FIELD SCORING'
WRITE(6,*) '      4. FORCE FIELD SCORING ONLY'
WRITE(6,*) 'Default is 3'
CALL INTIN('VERSN',3,1,4,VERSN)
WRITE(1, '(I2)') VERSN

WRITE(6,*)
WRITE(6,*) 'SINGLE (1) or SEARCH (2) mode? [2] '
READ(5, '(I1)') INT1
IF(INT1.EQ.1) THEN
    MODE = 'SINGLE'
    WRITE(1,3) 'SINGLE'
ELSE
    MODE = 'SEARCH'
    WRITE(1,3) 'SEARCH'
ENDIF

WRITE(6,*)
WRITE(6,*) 'Name of receptor sphere cluster file: '
100 CONTINUE
WRITE(6, '(A9,$)') ' CLUFIL> '
READ(5,9) CLUFIL
IF (CLUFIL .EQ. ' ') GOTO 100
WRITE(1,9) CLUFIL

WRITE(6,*)
WRITE(6,*) 'Enter numbers of clusters to use,'
WRITE(6,*) 'separated by spaces [1]: '
WRITE(6, '(A7,$)') ' LINE> '
READ(5,9) CLULINE
IF (CLULINE .EQ. ' ') CLULINE = '1'
WRITE(1,9) CLULINE

IF (VERSN .NE. 4) THEN
    WRITE(6,*)
    WRITE(6,*) 'Name of file containing CONTACT scoring grid: '

```

```

120     CONTINUE
        WRITE(6, '(A9,$)') ' MAPNAM> '
        READ(5,9)MAPNAM
        IF (MAPNAM .EQ. ' ') GOTO 120
        WRITE(1,9)MAPNAM
        ENDIF

c      Delphi, Force Field values get written at the end of INDOCK.

        IF (VERSN .EQ. 2) THEN
            WRITE(6,*)
            WRITE(6,*)'Name of file with DELPHI electrostatic map: '
140        CONTINUE
            WRITE(6, '(A9,$)') ' PHIFIL> '
            READ(5,9)PHIFIL
            IF (PHIFIL .EQ. ' ') GOTO 140
            ENDIF

        IF (VERSN .EQ. 3 .OR. VERSN .EQ. 4) THEN
            WRITE(6,*)
            WRITE(6,*)'Name of file with van der Waals parameters: '
160        CONTINUE
            WRITE(6, '(A9,$)') ' VDWFIL> '
            READ(5,9)VDWFIL
            IF (VDWFIL .EQ. ' ') GOTO 160
            WRITE(6,*)
            WRITE(6, '(A,A)') ' Prefix name for force field grid ',
&          'files from CHEMGRID: '
180        CONTINUE
            WRITE(6, '(A9,$)') ' PREFIX> '
            READ(5,9)PREFIX
            IF (PREFIX .EQ. ' ') GOTO 180

            WRITE(6, '(A,A)') ' Interpolate force field scores between',
&          ' grid points? [y] or n'
            WRITE(6, '(A8,$)') ' INTRP> '
            READ(5,2) DUM1
            IF (DUM1 .EQ. 'y' .OR. DUM1 .EQ. 'Y' .OR.
&          DUM1 .EQ. ' ') THEN
                INTRP = 1
            ELSE
                INTRP = 0
            ENDIF

            ESCALE = 1.0
            VSCALE = 1.0
            VDWMAX = 1.0E10
            WRITE(6,*) 'Change vdwmax, escale, or vscale values '
            WRITE(6,*) 'affecting the force field scoring? y or [n]'
            READ(5,2) DUM1
            IF (DUM1 .EQ. 'y' .OR. DUM1 .EQ. 'Y') THEN
                WRITE(6,*) 'Maximum vdw energy a ligand atom can receive'
&          ', in kcal/mol [1.0E10]'
                WRITE(6, '(A9,$)') ' VDWMAX> '
                READ(5,9) CHARIN
                IF (CHARIN .NE. ' ') THEN
                    READ(CHARIN,*) VDWMAX
                ENDIF
            ENDIF

```

```

        WRITE(6,*) 'Electrostatic scaling factor [1.0]'
        WRITE(6, '(A9,$)') ' ESCALE> '
        READ(5,9) CHARIN
        IF (CHARIN .NE. ' ') THEN
            READ(CHARIN,*) ESCALE
        ENDIF
        WRITE(6,*) 'van der Waals scaling factor [1.0]'
        WRITE(6, '(A9,$)') ' VSCALE> '
        READ(5,9) CHARIN
        IF (CHARIN .NE. ' ') THEN
            READ(CHARIN,*) VSCALE
        ENDIF
    ENDIF
ENDIF

WRITE(6,*)
WRITE(6,*) 'Enter dislim: [1.5]'
CALL REALIN('DISLIM',1.5,0.0,10000.,DISLIM)
WRITE(6,*)

WRITE(6,*) 'Enter number of pairs in a match [4]: '
CALL INTIN('NODLIM',4,4,100,NODLIM)
WRITE(6,*)

WRITE(6,*) 'Enter minimum number of pairs in a ',
& 'match [4]: '
CALL INTIN('LOWNOD',4,4,100,LOWNOD)
IF (LOWNOD.GT.NODLIM) LOWNOD=NODLIM
WRITE(6,*)

WRITE(6,*) 'Enter ratiom [0.0]: '
CALL REALIN('RATIOM',0.0,0.0,1.0,RATIOM)

WRITE(1,*) DISLIM,NODLIM,RATIOM,LOWNOD

WRITE(6,*)
WRITE(6,*)
WRITE(6,*) 'BIN PARAMETERS'
WRITE(6,*) '-----'
WRITE(6,*) 'Enter ligand bin size [1.0]: '
CALL REALIN('LBINSZ',1.0,0.0,100000.,LBINSZ)
IF (LBINSZ .EQ. 0.0) LBINSZ = 1.0
WRITE(6,*)

WRITE(6,*) 'Enter ligand bin overlap [0.0]: '
CALL REALIN('LOVLAP',0.0,0.0,100000.,LOVLAP)
WRITE(6,*)

WRITE(6,*) 'Enter receptor bin size [1.0]: '
CALL REALIN('SBINSZ',1.0,0.0,100000.,SBINSZ)
IF (SBINSZ .EQ. 0.0) SBINSZ = 1.0
WRITE(6,*)

WRITE(6,*) 'Enter receptor bin overlap [0.0]: '
CALL REALIN('SOVLAP',0.0,0.0,100000.,SOVLAP)

WRITE(1,8) LBINSZ,LOVLAP,SBINSZ,SOVLAP

```

```

        WRITE(6,*)
        WRITE(6,*) 'Name of file containing ligand(s) coordinates:'
200 CONTINUE
        WRITE(6, '(A9,$)') ' LIGFIL> '
        READ(5,9) LIGFIL
        IF (LIGFIL .EQ. ' ') GOTO 200
        WRITE(1,9) LIGFIL

        WRITE(6,*)
        WRITE(6,*) 'Name of output file:'
220 CONTINUE
        WRITE(6, '(A9,$)') ' OUTFIL> '
        READ(5,9) OUTFIL
        IF (OUTFIL .EQ. ' ') GOTO 220
        WRITE(1,9) OUTFIL

        WRITE(6,*)
        WRITE(6,*) 'Number of allowed bad contacts for an',
& ' orientation [0]:'
        CALL INTIN('ICTBMP',0,0,100000,ICTBMP)
        WRITE(1,5) ICTBMP

        IF(MODE.EQ.'SEARCH') THEN
            WRITE(6,*)
            WRITE(6,*)
            WRITE(6,*) 'SEARCH MODE OPTIONS'
            WRITE(6,*) '-----'
            WRITE(6, '(A,A)') 'Enter minimum number of non-Hydrogen atoms'
& ', per ligand [1]: '
            CALL INTIN('NATMIN',1,1,100000,NATMIN)
            WRITE(6,*)

            WRITE(6, '(A,A)') 'Enter maximum number of non-Hydrogen ',
& 'atoms per ligand [40]:'
            CALL INTIN('NATMAX',40,1,100000,NATMAX)
            WRITE(6,*)

            WRITE(6,*) 'Number of ligands to save [50]: '
            CALL INTIN('NSAV',50,0,100000,NSAV)
            WRITE(6,*)

            WRITE(6,*) 'Number of ligands to save with per atom ',
& 'scoring [0]: '
            CALL INTIN('NNSAV',0,0,100000,NNSAV)

            WRITE(1,*) NATMIN,NATMAX,NSAV,NNSAV

            WRITE(6,*)
            WRITE(6,*)
            WRITE(6,*) 'RESTART OPTIONS'
            WRITE(6,*) '-----'
            WRITE(6,*) 'For a restart run, enter 1 [0]:'
            CALL INTIN('IRESTR',0,0,1,IRESTR)
            WRITE(6,*)

            WRITE(6,*) 'Enter total number of molecules to search [1000]:'
            CALL INTIN('MOLTOT',1000,1,10000000,MOLTOT)
            WRITE(6,*)

```

```

WRITE(6,*) 'Enter number of molecules to dock before ',
& 'saving [100]: '
CALL INTIN('MOLSAV',100,1,10000000,MOLSAV)
IF (MOLSAV .GT. MOLTOT) MOLSAV = MOLTOT

```

```

WRITE(6,*) 'Enter number of molecules to skip at the ',
& 'beginning of the database [0]: '
CALL INTIN('MOLSKP',0,0,1000000,MOLSKP)

```

```

WRITE(1,*) IRESTR,MOLTOT,MOLSAV,MOLSKP

```

```

WRITE(6,*)
WRITE(6,*) 'Write orientations with hydrogens? [y] or n'
WRITE(6, '(A9,$)') ' INCHYD> '
READ(5, '(A1)') INCHYD
IF (INCHYD .EQ. 'n' ) THEN
    INCHYD = 'N'
ELSEIF (INCHYD .NE. 'N') THEN
    INCHYD = 'Y'
ENDIF
WRITE(1, '(A1)') INCHYD
ENDIF

```

```

IF (MODE.EQ. 'SINGLE') THEN
    WRITE(6,*)
    WRITE(6,*)
    WRITE(6,*) 'SINGLE MODE OPTIONS'
    WRITE(6,*) '-----'
    WRITE(6,*)
    WRITE(6,*) 'To use ligand SPHERES, enter S:[C] '
    WRITE(6, '(A9,$)') ' LIGTYP> '
    READ(5, '(A1)') ligtyp
    IF (LIGTYP .EQ. 's' .OR. LIGTYP .EQ. 'S') THEN
        LIGTYP = 'S'
    ELSE
        LIGTYP = 'C'
    ENDIF
    WRITE(1, '(A1)') ligtyp
    IF (LIGTYP.EQ. 'S') THEN

```

240

```

        WRITE(6,*)
        WRITE(6,*) 'Name of ligand sphere file: '
        CONTINUE
        WRITE(6, '(A9,$)') ' LGSFIL> '
        READ(5,9) LGSFIL
        IF (LGSFIL .EQ. ' ') GOTO 240
    WRITE(1,9) LGSFIL
    ENDIF

```

```

WRITE(6,*)
IF (VERSN .EQ. 1 .OR. VERSN .EQ. 2 .OR. VERSN .EQ. 3) THEN
    WRITE(6,*) 'Minimum CONTACT score to write an ',
& 'orientation [100]:'
    CALL REALIN('MINSCR',100.,-999.,100000.,MINSCR)
    WRITE(1, '(F9.4)') MINSCR
    WRITE(6,*)
ENDIF

```

```

WRITE(6,*)
IF (VERSN .EQ. 2 ) THEN
  WRITE(6,*) 'Maximum DELPHI electrostatic score to ',
&      'write an orientation [1000000]:'
  CALL REALIN('MAXSCR',1000000.,-100000.,100000000.,MAXSCR)
  WRITE(1, '(F10.1)') MAXSCR
  WRITE(6,*)
ELSEIF (VERSN .EQ. 3 .OR. VERSN .EQ. 4 ) THEN
  WRITE(6,*) 'Maximum FORCE FIELD score to write an ',
&      'orientation [1000000]:'
  CALL REALIN('MAXSCR',1000000.,-100000.,100000000.,MAXSCR)
  WRITE(1, '(F10.1)') MAXSCR
  WRITE(6,*)
ENDIF

c      NOTE: if -1 is entered, RMS scoring will NOT be done.

WRITE(6,*) 'Calculate rms deviations from a fixed ',
&      'orientation? y or [n] '
READ(5, '(A1)') QRMSD
IF (QRMSD.EQ. 'y' .OR. QRMSD .EQ. 'Y') THEN
  WRITE(6,*)
  WRITE(6,*) 'Enter maximum allowed rms deviation [5.0]: '
  CALL REALIN('USRDEV',5.0,0.0,1000.,USRDEV)
ELSE
  USRDEV = -1.0
ENDIF
WRITE(1, '(F9.4)') USRDEV
ENDIF

WRITE(6,*)
WRITE(6,*) 'Use zooming? y or [n] '
READ(5, '(A1)') DUM1
IF (DUM1 .EQ. 'y' .OR. DUM1 .EQ. 'Y') THEN
  WRITE(6,*)
  WRITE(6,*) 'Enter number of bin expansions to perform [1]: '
  CALL INTIN('EXPMAX',1,1,1000,EXPMAX)
  WRITE(1, '(I3)') EXPMAX
  WRITE(6,*)
  WRITE(6,*) 'Enter allowed contacts for scaling [0]: '
  CALL INTIN('FCTBMP',0,0,1000,FCTBMP)
  WRITE(1, '(I3)') FCTBMP
ELSE
  WRITE(1, '(I3)') 0
ENDIF

c      write dock3 specific scoring info here

IF (VERSN .EQ. 2) THEN
  WRITE(1,9) PHIFIL
ELSEIF (VERSN .EQ. 3 .OR. VERSN .EQ. 4) THEN
  WRITE(1,9) VDWFIL
  WRITE(1,9) PREFIX
  WRITE(1,5) INTRP
  WRITE(1,*) VDWMAX, ESCALE, VSCALE
ENDIF
CLOSE(1)
WRITE(6,*)

```

```
WRITE(6,*) 'INDOCK COMPLETED.'  
END
```

C-----

```
SUBROUTINE REALIN(PROMPT,DEFAULT,LOWER,UPPER,PARAM)
```

```
c      Prompts for and reads real input.  
c      Replaces illegal values with the appropriate default.
```

```
CHARACTER*6 PROMPT  
CHARACTER*80 CHARIN,DESCRP  
REAL DEFAULT,LOWER,UPPER,PARAM
```

```
110  FORMAT(A,' [,F4.1,']')  
120  FORMAT(1x,A6,'> ',)$  
130  FORMAT(A80)
```

```
WRITE(6,120) PROMPT  
READ(5,130) CHARIN  
IF (CHARIN .EQ. ' ') THEN  
    PARAM = DEFAULT  
ELSE  
    READ(CHARIN,*) PARAM  
    IF (PARAM .LT. LOWER .OR. PARAM .GT. UPPER) THEN  
        PARAM = DEFAULT  
    ENDIF  
ENDIF  
RETURN  
END
```

C-----

```
SUBROUTINE INTIN(PROMPT,DEFAULT,LOWER,UPPER,PARAM)
```

```
c      Prompts for and reads integer input.  
c      Replaces illegal values with the appropriate default.
```

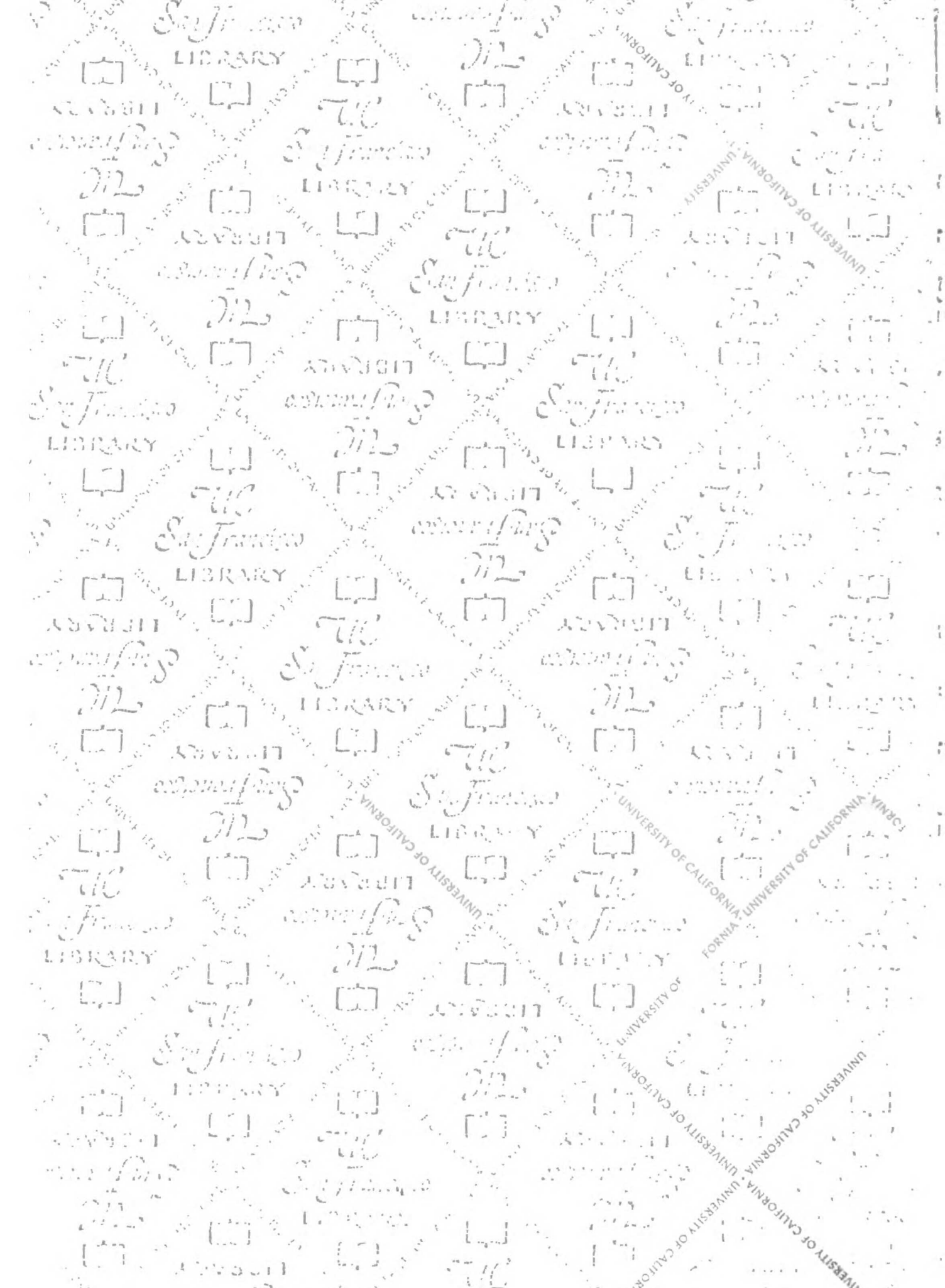
```
CHARACTER*6 PROMPT  
CHARACTER*80 CHARIN,DESCRP  
INTEGER DEFAULT,LOWER,UPPER,PARAM
```

```
110  FORMAT(A,' [,I2,']')  
120  FORMAT(1x,A6,'> ',)$  
130  FORMAT(A80)
```

```
WRITE(6,120) PROMPT  
READ(5,130) CHARIN  
IF (CHARIN .EQ. ' ') THEN  
    PARAM = DEFAULT  
ELSE  
    READ(CHARIN,*) PARAM  
    IF (PARAM .LT. LOWER .OR. PARAM .GT. UPPER) THEN  
        PARAM = DEFAULT  
    ENDIF  
ENDIF  
RETURN  
END
```


References

1. Shoichet, B. K., Bodian, D. L. and Kuntz, I. D., "Molecular docking using shape descriptors." *J Comp Chem* 13 380-397 (1992).
2. Meng, E. C., Shoichet, B. K. and Kuntz, I. D., "Automated docking with grid-based energy evaluation." *J Comp Chem* 13 505-524 (1992).



LIBRARY
USE ONLY

