Computational Methods for Protein Folding: Advances in Chemical Physics, Volume 120.
Edited by Richard A. Friesner. Series Editors: I. Prigogine and Stuart A. Rice.

ISBNs: 0-471-20955-4 (Hardback); 0-471-22442-1 (Electronic)

COMPUTATIONAL METHODS FOR PROTEIN FOLDING

A SPECIAL VOLUME OF ADVANCES IN CHEMICAL PHYSICS

VOLUME 120

COMPUTATIONAL METHODS FOR PROTEIN FOLDING

ADVANCES IN CHEMICAL PHYSICS
VOLUME 120

Edited by

RICHARD A. FRIESNER

Columbia University, New York, NY

Series Editors

I. PRIGOGINE

Center for Studies in Statistical Mechanics
and Complex Systems
The University of Texas
Austin, Texas
and
International Solvay Institutes
Universite Libre de Bruxelles
Brussels, Belgium

and

STUART A. RICE

Department of Chemistry
and
The James Franck Institute
The University of Chicago
Chicago, Illinois

AN INTERSCIENCE PUBLICATION
A JOHN WILEY & SONS, INC. PUBLICATION

ISBN 0-471-22442-1

This title is also available in print as ISBN 0-471-20955-4.

For more information about Wiley products, visit our web site at www.Wiley.com.

CONTRIBUTORS TO VOLUME 120

BENOIT CROMP, Département de Chimie, Université de Montréal, Montréal, Québec, Canada; Centre de Recherche en Calcul Appliqué, Montréal, Québec, Canada; and Protein Engineering Network of Centers of Excellence, Edmonton, Alberta, Canada

R. I. DIMA, Institute for Physical Science and Technology and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD, U.S.A.

AARON R. DINNER, New Chemistry Laboratory, University of Oxford, Oxford, U.K.

RON ELBER, Department of Computer Science, Cornell University, Ithaca, NY, U.S.A.

VOLKER A. EYRICH, Department of Chemistry and Center for Biomolecular Simulation, Columbia University, New York, NY, U.S.A.

ANTHONY K. FELTS, Department of Chemistry, Rutgers University, Wright-Rieman Laboratories, Piscataway, NJ, U.S.A.

CHRISTODOULOS A. FLOUDAS, Department of Chemical Engineering, Princeton University, Princeton, NJ, U.S.A.

RICHARD A. FRIESNER, Department of Chemistry and Center for Biomolecular Simulation, Columbia University, New York, NY, U.S.A.

EMILIO GALLICCHIO, Department of Chemistry, Rutgers University, Wright-Rieman Laboratories, Piscataway, NJ, U.S.A.

JOHN GUNN, Schrödinger, Inc., New York, NY, U.S.A.; Centre de Recherche en Calcul Appliqué, Montréal, Québec, Canada; and Protein Engineering Network of Centers of Excellence, Edmonton, Alberta, Canada

PIERRE-JEAN L'HEUREUX, Département de Chimie, Université de Montréal, Montréal, Québec, Canada; Centre de Recherche en Calcul Appliqué, Montréal, Québec, Canada; and Protein Engineering Network of Centers of Excellence, Edmonton, Alberta, Canada

MARTIN KARPLUS, New Chemistry Laboratory University of Oxford, Oxford, U.K.; Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, U.S.A.; and Laboratoire de Chimie Biophysique, Institut le Bel, Université Louis Pasteur, Strasbourg, France

JOHN L. KLEPEIS, Department of Chemical Engineering, Princeton University, Princeton, NJ, U.S.A.

D. K. KLIMOV, Institute for Physical Science and Technology and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD, U.S.A.

ANDRZEJ KOLINSKI, Laboratory of Computational Genomics, Danforth Plant Science Center, Creve Coeur, MO, U.S.A.; and Department of Chemistry, University of Warsaw, Warsaw, Poland

RONALD M. LEVY, Department of Chemistry, Rutgers University, Wright-Rieman Laboratories, Piscataway, NJ, U.S.A.

ÉRIC MARTINEAU, Département de Chimie, Université de Montréal, Montréal, Québec, Canada; Centre de Recherche en Calcul Appliqué, Montréal, Québec, Canada; and Protein Engineering Network of Centers of Excellence, Edmonton, Alberta, Canada

JAROSLAW MELLER, Department of Computer Science, Cornell University, Ithaca, NY, U.S.A.; and Department of Computer Methods, Nicholas Copernicus University, Torun, Poland

HEATHER D. SCHAFROTH, Department of Chemical Engineering, Princeton University, Princeton, NJ, U.S.A.

JEFFREY SKOLNICK, Laboratory of Computational Genomics, Danforth Plant Science Center, Creve Coeur, MO, U.S.A.

SUNG-SAU SO, Hoffman-La Roche, Inc., Discovery Chemistry, Nutley, NJ, U.S.A.

DARON M. STANDLEY, Schrödinger Inc., New York, NY, U.S.A.

D. THIRUMALAI, Institute for Physical Science and Technology and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD, U.S.A.

ANDERS WALLQVIST, Department of Chemistry, Rutgers University, Wright-Rieman Laboratories, Piscataway, NJ, U.S.A.

KARL M. WESTERBERG, Department of Chemical Engineering, Princton University, Princeton, NJ, U.S.A.

INTRODUCTION

Few of us can any longer keep up with the flood of scientific literature, even in specialized subfields. Any attempt to do more and be broadly educated with respect to a large domain of science has the appearance of tilting at windmills. Yet the synthesis of ideas drawn from different subjects into new, powerful, general concepts is as valuable as ever, and the desire to remain educated persists in all scientists. This series, *Advances in Chemical Physics*, is devoted to helping the reader obtain general information about a wide variety of topics in chemical physics, a field that we interpret very broadly. Our intent is to have experts present comprehensive analyses of subjects of interest and to encourage the expression of individual points of view. We hope that this approach to the presentation of an overview of a subject will both stimulate new research and serve as a personalized learning text for beginners in a field.

I. PRIGOGINE
STUART A. RICE

PREFACE

The first attempts to model proteins on the computer began almost 30 years ago. Over the past three decades, our understanding of protein structure and dynamics has dramatically increased as a result of rapid advances in both theory and experiment. The Protein Data Bank (PDB) now contains more than 10,000 high-resolution protein structures. The human genome project and related efforts have generated an order of magnitude more protein sequences, for which we do not yet know the structure. Spectroscopic measurement techniques continue to increase in resolution and sensitivity, allowing a wealth of information to be obtained with regard to the kinetics of protein folding and unfolding, complementing the detailed structural picture of the folded state. In parallel to these efforts, algorithms, software, and computational hardware have progressed to the point where both structural and kinetic problems may be studied with a fair degree of realism.

Despite these advances, many major challenges remain in understanding protein folding at both a conceptual and practical level. There is still significant debate about the role of various underlying physical forces in stabilizing a unique native structure. Efforts to translate physical principles into practical protein structure prediction algorithms are still at an early stage; most successful prediction algorithms employ knowledge-based approaches that rely on examples of existing protein structures in the PDB, as well as on techniques of computer science and statistics. Theoretical modeling of the dynamics of protein folding faces additional difficulties; there is a much smaller body of experimental data, which is typically at relatively low resolution; carrying out computations over long time scales requires either very large amounts of computer time or the use of highly approximate models; and the use of statistical methods to analyze the data is still in its infancy.

The importance of the protein folding problem—underscored by the recent completion of the human genome sequence—has led to an explosion of theoretical work in areas of both protein structure prediction and kinetic modeling. An exceptionally wide variety of computational models and techniques are being applied to the problem, due in part to the participation of scientists from so many different disciplines: chemistry, physics, molecular biology, computer science, and statistics, to name a few. This has made the field very exciting for those of us working in it, but it also poses a challenge; how can the key issues in state of the art research be communicated to different audiences, given the interdisciplinary nature of the task at hand and the methods being brought to bear on it?

The objective of this volume of *Advances in Chemical Physics* is to discuss recent advances in the computational modeling of protein folding for an audience of physicists, chemists, and chemical physicists. Many of the contributors to this volume have their roots in chemical physics but have committed a significant fraction of their resources to studying biological systems. The chapters thus address the target audience but incorporate approaches from other areas because they are relevant to the methods that the various authors have developed in their laboratories. While some of the chapters contain review sections, the principal focus is on the authors' own research and recent results.

When modeling protein folding the key questions are (a) the nature of the physical model to be used and (b) the questions that the calculations are aimed at answering. It is impossible in a single volume to cover all of the different approaches that are currently being used in research on protein folding. Nevertheless, a reasonably broad spectrum of computational methods is represented here, as is briefly described below. The volume is organized so as to group together contributions in which similar approaches are adopted.

The simplest models of proteins involve representations of the amino acids as beads on a chain (typically taken to be hydrophobic or hydrophilic, depending upon the identity of the amino acid) embedded in a lattice. Primitive models of this type employ a simple lattice such as a cubic lattice, and they use a single center to represent each amino acid. These models are very fast computationally, but lack a level of detail (both structurally and in their potential energy function) to permit prediction of protein structure from the amino acid sequence. On the other hand, they can be extremely valuable in providing conceptual insight into the general thermodynamic and kinetic issues as to why and how proteins fold into a unique native state; they can also be profitably used to model folding kinetics, as well as to make testable predictions for such kinetics that can be compared with experimental data. The contributions of Thirumulai et al. and Dinner et al. discuss models of this type, presenting both conceptual insights into the basis of protein folding and results for modeling of specific protein folding events.

Reduced models of proteins (i.e., models not containing complete atomic detail) can be used to make structural predictions, either by allowing assessment of the fitness of a protein structure already in the PDB as a model for an unknown sequence (''threading'') or by carrying out Monte Carlo simulations using the model and a suitable potential energy function. The contribution by Meller and Elber describes a classical threading approach in which the amino acid sequence is ''threaded'' in an optimal fashion onto a set of candidate template structures using dynamic programming techniques, and the suitability of the template is evaluated by a potential energy function. These authors have worked out new methods for optimizing such functions, which are discussed in detail in their chapter.

If a reduced (or other) model is used to predict protein structure via simulation, without direct reference to structures in the PDB, this is referred to as ''*ab initio* protein'' structure prediction. Potential energy functions for *ab initio* prediction can be derived either from physical chemical principles or from a ''knowledge-based'' approach based on statistics from the PDB (e.g., the probability of observing a residue–residue distance for a given pair of amino acids). For reduced models, the use of knowledge-based potential of some sort is mandated. The contributions of Eyrich et al., Skolnick and Kolinsiki, and L'Heureux et al. derive originally from an *ab initio* approach using reduced models. However, all of these groups have in the past several years increasingly incorporated empirical elements from threading and other such approaches, so that what is described in these contributions is more of an attempt to integrate reduced model simulations with additional information and techniques that can improve practical structure prediction results. Several of these research groups have entered the CASP (Critical Assessment of Protein Structure Prediction) blind test experiments, which allow a comparative evaluation of the prediction accuracy of the different methods employed by the participants; results from the most recent such experiment, CASP4 (not reported in this volume because the results were available subsequent to submission of most of the chapters), were encouraging with regard to the ability of these hybrid methods to provide improvement in many cases over methods not incorporating simulations.

The use of models employing an atomic level of detail (e.g. a molecular mechanics potential function) in addressing the protein folding problem presents significant difficulties for two reasons: (1) A large expenditure of computation time is required to evaluate the model energy at each configuration; (2) the quality of the potential energy functions and solvation model are critical in being able to accurate compare the stability of alternative structures. The contribution by Klepeis et al. discusses both algorithms designed to reduce the required computational effort by sampling phase space more efficiently and a wide variety of applications of atomic level models using these more efficient sampling techniques. The contribution from Wallqvist et al. is more narrowly focused on a single problem: the use of detailed atomic potential functions in conjunction with a continuum solvation model to distinguish native and ''native-like'' protein structures from ''decoys''—alternative structures generated by various means and intended to challenge the model's accuracy. Both of these contributions demonstrate that considerable progress is being made in the application of atomic level models with regard to improving both accuracy and efficiency.

In the end, a thorough description of all aspects of protein folding will require the use of the full range of models and methods discussed in this volume. In the simplest hierarchical picture, one can imagine using inexpensive reduced models to generate low-resolution structures that can then be refined

using more detailed (and computationally expensive) approaches. Although progress will undoubtedly continue in the development of physical chemical models, empirical information and phenomenological approaches will always provide additional speed and reliability if practical results are desired. How to best combine all of these elements represents one of the principal issues facing those working in the field; it also exemplifies the need for new ideas and approaches.

Columbia University
New York, New York

RICHARD A. FRIESNER

CONTENTS

Statistical Analysis of Protein Folding Kinetics 1
By Aaron R. Dinner, Sung-Sau-So, and Martin Karplus

Insights into Specific Problems in Protein Folding Using Simple Concepts 35
By D. Thirumalai, D. K. Klimov, and R. I. Dima

Protein Recognition by Sequence-to-Structure Fitness: Bridging Efficiency and Capacity of Threading Models 77
By Jaroslaw Meller and Ron Elber

A Unified Approach to the Prediction of Protein Structure and Function 131
By Jefferey Skolnick and Andrzej Kolinski

Knowledge-Based Prediction of Protein Tertiary Structure 193
By Pierre-Jean L'Heureux, Benoit Cromp, Éric Martineau, and John Gunn

Ab Initio Protein Structure Prediction Using a Size-Dependent Tertiary Folding Potential 223
By Volker A. Eyrich, Daron M. Standley, and Richard A. Friesner

Deterministic Global Optimization and *Ab Initio* Approaches for the Structure Prediction of Polypeptides, Dynamics of Protein Folding, and Protein–Protein Interactions 265
By John L. Klepeis, Heather D. Schafroth, Karl M. Westerberg, and Christodouls A. Floudas

Detecting Native Protein Folds Among Large Decoy Sites with the OPLS All-Atom Potential and the Surface Generalized Born Solvent Model 459
By Anders Wallqvist, Emilio Gallicchio, Anthony K. Felts, and Ronald M. Levy

Author Index 487

Subject Index 507

SUBJECT INDEX

Ab initio prediction:
  native *vs.* decoy protein conformations, decoy data sets, 465
  polypeptide structure prediction, future issues, 360–362
  sequence-structure-function paradigm, 138–143, 147–148
    biochemical function, 174–175
    distance geometry, 160
    future issues, 180–186
    generalized comparative modeling, 155–156
    methodology, 165–167
    test protein results, 167–169
    unified folding method, 147–148
  tertiary protein structure:
    folding potential:
      PDB-derived secondary structure simulations, 238–246
        methodology, 238–241
        size-dependent potential, 241–245
        terminal loop definition and truncation, 245
        three-dimensional topologies, 245–246
      predicted-derived secondary structure simulations, 246–260
      research background, 224–226
      size-dependent potential energy function, 226–237
        error identification, 226–230
        improvements in, 230–233
        results, 233–237
    knowledge-based approach *vs.,* 195–197
Active site identification, sequence-structure-function prediction, 172–174
Adenosine triphosphate (ATP), chaperonin-facilitated protein folding mechanism, 64–66
  unfolding activity, 66–67
Alanine, oligopeptide structure prediction, computational studies, 303–312
All-atom force-fields, native *vs.* decoy protein conformation, research background, 460–462
αBB algorithm:
  deterministic global optimization:
    parallelization issues, 363
    twice continuously differentiable NLPs, 270, 276–279
      general nonconvex terms, 272–274
      geometrical representation, 282–287
      procedures and solutions, 276–279
      solution enclosure, 279–282
  oligopeptide structure prediction:
    free energy modeling, 317–318
    global optimization, 298–300
  protein folding:
    coil-to-helix transition, 371
    global optimization approach, 367–369
    minimum energy conformation, 393
  sparse restraints, structural refinement, 345–347
Alpha carbon representation, sequence-structure-function paradigm, *ab initio* prediction, 138–143
Alpha-helical ground state, protein folding, coil-to-helix transition, reaction coordinates, 383–386
AMBER all-atom force field model, native protein detection, 461
Ambiguous constraints, tertiary protein structure:
  implementation, 205
  results, 208–209
Amino acid substitutions, protein-protein interactions, binding site structure prediction, 416–418
Anfinsen's hypothesis:
  chaperonin-facilitated protein folding mechanism, 63–66

Anfinsen's hypothesis: (*Continued*)
oligopeptide structure prediction, free energy modeling, 312–314
protein folding mechanisms, 267–269
Angle constraints, tertiary protein structure:
implementation and results, 207–208
knowledge-based prediction, 202
ARAS program, protein-protein interactions, binding site structure prediction, 420–421
ASP sets, oligopeptide structure prediction:
computational studies, 301–312
global optimization framework, 296–300
solvation energy models, 293–296
ASTRO-FOLD approach, polypeptide structure prediction, 360–362
Atomic model reconstruction, sequence-structure-function prediction, 160–161
reduced/atomic model compatibility:
atomic reconstruction, 170–171
structural refinement feasibility, 171–172
structural reproducibility, 169–170

Barnase, disulfide bond engineering, 62–63
β-hairpin fold, ECEPP/3 prediction, 403–405
β-strand-based superstructure, polypeptide structure prediction, 361–362
β-strand pairing, tertiary protein structure, local templates, 217–218
Bezier splines, tertiary protein structure, *ab initio* folding potential predictions, 229–230
Binding affinity prediction, protein-protein interactions, 408–414, 430–442
computational studies, 440–442
decomposition techniques, 431
flexible models, 411–414
minimization scheme, 438–440
modeling, 431–438
rigid models, 410–411
Binding site structure prediction, protein-protein interactions, 407–408
computational studies, 421–430
crystallographic data, 416–418
computational studies, 421–424
HLD-DR3 binding sites, 424–427
murine structure binding sites, 427–430
deterministic global optimization, 420–421
histocompatibility molecules, 415
modeling techniques, 418–420
Biochemical function, sequence-structure-function prediction, 172–178
ab initio folding, 174–175
active site identification, 172–174
future issues, 183–184
threaded structures, 175–178
BLAST profiling, sequence-structure-function prediction, Fischer database applications, 151
BLOCKS technique, sequence-structure-function prediction:
biochemical function prediction, 175–178
future issues, 183–184
BLOSUM matrix, sequence-structure-function prediction, first-pass threading, 148–149
Born-Oppenheimer approximation, oligopeptide structure prediction, potential energy models, 288–291
Bovine pancreatic trypsin inhibitor (BPTI):
disulfide bonds in protein folding, 53–55
modeling techniques, 58–62
protein folding kinetics, 56–58
BPMPD program, energy parameter optimization, 89
Branch-and-bound strategies:
deterministic global optimization, 363
sequence-structure-function prediction, threading procedures, 136–137

Calcium-binding protein, tertiary protein structure, distance constraints, 206–207
Cartesian molecular dynamics, sparse restraints, structural refinement, 342–345
Cartesian root mean square deviation (cRMSD), protein-protein interactions, binding site structure prediction, 423–430
Catchment regions, protein folding, potential energy surfaces, 367
Chaperonin-facilitated protein folding mechanism, 63–69
GroEL unfolding, 66–68
stretching, unfolding by, 68–69
CHARMM force field:
native protein detection, 461
solvation free energy model, 483–484
potential energy surface, 404–405
sequence-structure-function paradigm, reduced/atomic model compatibility, 171–172

CLUSTALW database, sequence-structure-function prediction, first-pass threading, 148–149
Clustering algorithm, protein folding, 396–397
Coil-to-helix transitions, protein folding dynamics, 369–390
  pathways, 373–378
  rate disconnectivity graphs, 378–380
  reaction coordinates, 383–386
  solvated tetra-alanine, 386–390
  time evolution of quantities, 380–382
  transition rates and master equation, 371–373
  unsolvated tetra-alanine, 370–371
Collapse time, two-state protein folding, 50–51
Comparative modeling methods, sequence-structure-function prediction, 134, 155–165
  atomic model reconstruction, 160–161
  average model construction, 160
  Fischer database results, 161–164
  lattice chain, starting construction, 158
  lattice folding optimization, 158–159
  lattice model and force field, 157–158
  Modeller tool, 164–165
Competing basins of attraction (CBA), protein folding mechanisms, 52–53
Compstatin, sparse restraints, structural refinement:
  computational analysis, 347–353
  global optimization, torsion angle dynamics, 357–359
Computational modeling:
  oligopeptide structure prediction, 301–312
  protein-protein interactions, binding site structure prediction, 421–430
    crystallographic data, 421–424
    HLA-DR3 binding site prediction, 424–427
    HLA-DR1 pockets, 440–442
    murine binding site prediction, 427–430
  sparse restraints, structural refinement, 347–353
  tertiary protein structure, 197–198
Conformational Space Annealinig (CSA) algorithm, protein folding:
  minimum energy conformation, 393–394
  stationary points searching, 365–366
Conformational space reduction:
  protein folding, 43–44
  tertiary protein structure, 212
Constraint methods, tertiary protein structure:
  knowledge-based prediction, 201–210
    ambiguous constraints, 205, 208–209
    angle constraints, 202, 207–208
    distance constraints, 202, 206–207
    implementation, 204–205
    predictions, derivation from, 203–204, 209–210
  refinement, 219
Contact energies, protein folding kinetics, lattice models, 39
Control sets, energy parameter optimization, 87–88
Convexification, deterministic global optimization, twice continuously differentiable NLPs, feasible region convexification, 274
Convex lower bounding:
  deterministic global optimization, twice continuously differentiable NLPs, 275
    αBB algorithm, 277–279
  oligopeptide structure prediction, local minimum energy conformations, 318–321
Coriolis forces, sparse restraints, structural refinement, 342–345
Coulombic interactions:
  native *vs.* decoy protein conformations:
    dielectric model approximations, 478–480
    distance-dependent dielectric approximations, 480–481
    energy components, 476–478
    Park and Levitt decoys, 471–472
    SGB calculations, 463–464
  oligopeptide structure prediction, potential energy models, 290–291
Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiments:
  energy parameter optimization, 87–88
  native *vs.* decoy protein conformations:
    CASP3 targets, 474–475
    decoy data sets, 465
  predicted-derived secondary structure simulations:
    prediction contest results, 258–260
    tertiary protein structures, 247
  sequence-structure-function prediction:
    ab initio prediction, 140–143, 174–175
    comparative modeling techniques, 134

Critical Assessment of Techniques for Protein Structure Prediction (CASP) experiments: (*Continued*)
threading techniques, 135–137
tertiary folding simulation, PDB-derived and ideal structures, 238
size-dependent energy functions, 241–245
tertiary protein structure, knowledge-based prediction, 194–195
THOM2 model, self-recognition of folds, 118–120
Cross-validation, protein folding statistical analysis, 6
database elements, 11–16
two-descriptor models, 20–22
Crystallographic data, protein-protein interactions, binding site structure prediction, 421–424
Cubic splines, tertiary protein structure, *ab initio* folding potential predictions, 229–230
Cyclophilin A (CypA), chaperonin-facilitated protein folding mechanism, unfolding activity, 66–67
Cysteine residues, disulfide bonds in protein folding, bovine pancreatic trypsin inhibitor (BPTI), 53–55

DALI program:
energy parameter optimization, 87–88
THOM2 potential, self-recognition of folds, 120
Database properties, protein folding kinetics, 11–16
Data overfitting, protein folding statistical analysis, 6
Data sets, native *vs.* decoy protein conformations, 464–465
Decoy protein structures, detection *vs.* native conformations:
applications, 483–484
approximate effective dielectric models:
distance-dependent dielectric approximation, 480–481
screened Coulomb approximation, 478–480
CASP3 targets, 473–474
decoy data sets, 464–465
energy components, 474–478
force feld calculations, 462–464
Holm and Sander single decoys, 472–473
interior dielectric constant, 481–482
Park and Levitt decoys, 465–472
research background, 460–462
Decoy structures:
gap energy optimization, 101–107
THOM1 potential, native intermediates, 97–100
Deletions, gap energy structures, 107–108
DELPHI software package, protein-protein interactions, binding site structure prediction, 443–444
Dense sequence space, protein folding kinetics, 44–49
Designability issues, protein folding kinetics, dense sequence space, 47–48
Deterministic global optimization:
polypeptide structural prediction:
computational models, 301–312
free energy computational studies, 322–336
free energy modeling, 312–314
future issues, 360–362
global optimization framework, 296–300
harmonic approximation, 314–318
local minimum energy conformation, 318–322
parallelization, 363
potential energy models, 288–291
solvation energy models, 291–296
sparse restraint structural refinement, 336–359
algorithmic steps, 345–347
computational study, 347–353
DYANA *vs.* torsion angle dynamics (TAD), 353–356
energy modeling, 339–341
global optimization, 341
torsion angle dynamics (TAD), 341–345, 356–359
protein folding dynamics:
αBB algorithm, 367–369
coil-to-helix transitions, 369–390
pathways, 373–378
rate disconnectivity graphs, 378–380
reaction coordinates, 383–386
solvated tetra-alanine, 386–390
time evolution of quantities, 380–382
transition rates and master equation, 371–373
unsolvated tetra-alanine, 370–371

framework and implementations, 390–403
local stationary point search methods, 392–393
minima and transition rate search, 393–397
potential energy surface analysis, 397–403
future research, 403–405
potential energy surface, 366–367
secondary structure formation, 364
stationary points, 365–366
protein-protein interactions:
binding affinity prediction, 408–414, 430–442
computational studies, 440–442
decomposition techniques, 431
minimization scheme, 438–440
modeling, 431–438
binding site structure prediction, 407–408, 414–430
class II HLA molecule structures, 415
computational studies, 421–430
modeling techniques, 418–420
prediction techniques, 416–418
programming tools, 420–421
future research issues, 442–444
research background, 404–414
solution enclosure, 279–287
twice continuously differentiable NLPs, 269–279
αBB algorithm, 276–279
convex lower bounding, 275
feasible region convexification, 274–275
nonconvex terms, underestimation, 272–274
special structure, underestimation, 270–272
variable bound updates, 275–276
Dielectric models, native *vs.* decoy conformations:
distance-dependent dielectric approximation, 480–481
screened Coulomb approximation, 478–480
Diffusion-collision model, folding rates of proteins, 10–11
Dihedral angles, protein-protein interactions, binding site structure prediction, 418–420, 432–435
Distance constraints, tertiary protein structure:
implementation and results, 206–207
knowledge-based prediction, 202
Distance-dependent dielectric approximation, native *vs.* misfolded protein conformation, 480–481
Distance geometry, sequence-structure-function prediction, 160
Distance power-law potentials, pairwise models, 92–93
Distant-dependent pair energies, threading potentials, protein recognition, 79–80
Disulfide bonds, protein folding mechanism:
barnase engineering, 62–63
BPTI refolding, 53–55
modeling techniques, 58–62
proximity rule, 55–58
bovine pancreatic trypsin inhibitor (BPTI), 55
folding kinetics, 56–58
loop formation probability, 55
Double Z-score filter:
protein energy gaps, 112–114
THOM2 model, *vs.* pair energies, 120–126
THreading Onion Model 2 (THOM2), 114–126
HL test, 114–118
recognition excluded from training, 118–120
vs. pair energies, 120–126
DYANA protocol, sparse restraints, structural refinement, torsion angle dynamics (TAD) *vs.,* 353–356
Dynamic programming (DP) algorithm:
energy parameter optimization, 87–88
protein recognition, 79
sequence-structure-function prediction, threading procedures, 136–138
Dynamics analysis, protein folding dynamics, coil-to-helix transitions, 390–403

ECEPP/3 program:
oligopeptide structure prediction:
global optimization, 299–300
potential energy models, 290–291
solvation energy models, 294–296
protein folding mechanisms:
coil-to-helix transitions, 369–370
solvated tetra-alanine, 386–390
potential energy surfaces, 403–405
protein-protein interactions:
binding site structure prediction:

ECEPP/3 program: (*Continued*)
future issues, 442–444
global optimization, 438–440
solvation modeling, 438
global optimization, 420–421
sparse restraints, structural refinement:
αBB algorithm, 345–347
computational analysis, 348–353
DYANA protocol *vs.,* 354–356
energy modeling, 340–341
global optimization, torsion angle dynamics, 357–359
Effective Energy Function (EEF1), potential energy surface, 404–405
1eh2 proteins, *ab initio* simulations, 259
Eigenmode-following methods, protein folding:
coil-to-helix transition, 371–373
stationary points searching, 365, 392–393
uphill searches, 394–395
Energy-directed approach (EDA), oligopeptide structure prediction:
computational studies, 322–336
local minimum energy conformations, 319–321
Energy gaps:
double Z-score filter, 112–114
native *vs.* decoy protein conformations, 464
Holm and Sander single decoys, 472–474
screened Coulomb approximation, dielectric models, 478–480
optimization protocols, 101–107
protein folding kinetics, lattice models, 6–8
Z-score filter, 108–112
Energy minimization, oligopeptide structure prediction, global optimization framework, 297–300
Energy parameter optimization, protein recognition, 85–87
learning and control sets, 87–88
linear programming protocol, 88–89
Energy profile models:
distance power-law potentials, 92–93
minimal models, 91–92
new energy profiles, 93–100
parameter-free models, 90–91
protein recognition, 83–85
Entropic states, oligopeptide structure prediction:
free energy modeling, 313–314
harmonic approximation, 314–316
Equilibrium probability, protein folding, potential energy surface, 398–399
Escherichia coli, sequence-structure-function prediction:
future issues, 183–186
genome-scale threading, 153
ESMC methods, sequence-structure-function prediction, 144–146
Euler angles, protein-protein interactions, binding site structure prediction, 418–420
Excluded volume interactions, conformational space reduction, 43–44
Exhaustive enumeration, protein folding kinetics, lattice models, 41

Family protein recognition, THOM2 model *vs.* pair energies, 122–126
FASTA database, sequence-structure-function prediction, first-pass threading, 148–149
Fill algorithm, protein folding, 395–396
First-pass threading, sequence-structure-function prediction, 148–149
Fischer database, sequence-structure-function prediction:
generalized comparative modeling, 161–164
PROSPECTOR2 applications, 152
threading procedures, 149–151
Flavodoxin (2FX2), tertiary protein structure, ambiguous constraints, 208–209
Flexible-angle constraints, tertiary protein structure, constraint analysis, 203–204
Flexible models, protein-protein interactions, binding affinity prediction, 411–414
Folding rates of proteins:
statistical analysis, 9–26
database, 11–16
multiple-descriptor models, 19–24
physical observations, 24–26
review, 9–11
single-descriptor models, 16–19
tertiary protein structure, *ab initio* predictions:
PDB-derived secondary structure simulations, 238–246
methodology, 238–241
size-dependent potential, 241–245
terminal loop definition and truncation, 245
three-dimensional topologies, 245–246

predicted-derived secondary structure simulations, 246–260
research background, 224–226
size-dependent potential energy function, 226–237
error identification, 226–230
improvements in, 230–233
results, 233–237
Force field potentials:
native *vs.* decoy protein conformations, 462–464
sparse restraints, structural refinement, energy modeling, 339–341
Fragment lists:
protein-protein interactions, binding affinity prediction, 413–414
tertiary protein structure, search space limitation, local threading, 211–212
Fragment screening, tertiary protein structure, search space limitation, 212–213
Free energy directed approach (FEDA), oligopeptide structure prediction, local minimum energy conformations, 321
Free energy global minum (FEGM), oligopeptide structure prediction, computational studies, 322–336
Free energy modeling:
oligopeptide structure prediction, 312–318
computational studies, 322–336
protein folding:
coil-to-helix transition, 371–373
potential energy surface, 398
Frozen environment approximation (FEA):
sequence-structure-function prediction:
first-pass threading, 148–149
threading procedures, 136–138
THOM2 model, *vs.* pair energies, 120–126
Fuzzy functional forms (FFF), sequence-structure-function prediction:
active site identification, 173–174
biochemical function prediction, 175–178
future issues, 183–184

Gap energies:
deletions, 107–108
global alignments *vs.* local alignments, 108
optimization protocols, 101–107
training protocols, 103–107
Gapless threading:
distance power-law potentials, 92–93
energy parameter optimization, 86–87
learning and control sets, 87–88
pseudo-native structures, 104–107
Gaussian distribution, protein folding kinetics, random bond model, 40
GeneComp modeling, sequence-structure-function prediction, 183–186
General nonconvex terms, deterministic global optimization, twice continuously differentiable NLPs, 272–274
αBB algorithm, 276–279
Genetic algorithms:
protein folding statistical analysis, 5–6
three-descriptor models, 22–24
protein-protein interactions, binding affinity prediction, 413–414
sequence-structure-function prediction, *ab initio* prediction, 139–143
Genetic neural network (GNN) method, protein folding statistical analysis, lattice models, 8
Geometrical representation:
deterministic global optimization, twice continuously differentiable NLPs, 282–287
protein-protein interactions, binding site structure prediction, 418–420
tertiary protein structure, 198–199
knowledge-based prediction, 202–203
Gerschgorin's theorem, deterministic global optimization, twice continuously differentiable NLPs, 273–274
Gibbs-Boltzmann distribution, oligopeptide structure prediction, free energy modeling, 313–314
Global alignments, protein energy gaps, 108
double Z-score filter, 112–114
Global minimization. *See also* Deterministic global optimization
native protein detection, 461–462
oligopeptide structure prediction, 297–300
sparse restraints, structural refinement, 341
torsion angle dynamics (TAD), 356–359
twice continuously differentiable NLPs, solution enclosure, 280–282
Globular proteins:
sequence-structure-function paradigm, *ab initio* folding, 164–169

Globular proteins: (*Continued*)
tertiary protein structure, distance constraints, 206–207
GLO-DOCK software, protein-protein interactions, binding site structure prediction, 435–436
GLO-FOLD software, oligopeptide structure prediction, global optimization, 299–300
Glutathione, protein folding kinetics, 57–58
Go model:
protein folding kinetics, 40–41
two-state protein folding, 51
Gradient discontinuities, oligopeptide structure prediction, solvation energy models, 293–296
Graph-theory techniques, protein folding, coil-to-helix transition pathways, 374–378
Grid search algorithm, protein folding stationary points, 394
GroEL proteins:
chaperonin-facilitated protein folding mechanism, 64–66
stretching-induced unfolding, 68–69
unfolding activity, 66–67
disulfide bonds in protein folding, modeling techniques, 60–62
GroES proteins, chaperonin-facilitated protein folding mechanism, 64–65

Harmonic approximation, oligopeptide structure prediction, 314–316
potential energy models, 289–291
Helix pairing, tertiary protein structure:
constraint analysis, 203–204
fragment-based modeling, 213–214
Hessian matrix:
deterministic global optimization, twice continuously differentiable NLPs, 273–274
oligopeptide structure prediction, harmonic approximation, 315–316
protein folding:
global optimization, 369
potential energy surface, vibrational frequencies, 397–398
stationary points searching, 392–393
Hierarchical search algorithm, tertiary protein structure, 199–200
search space limitation, 212–213
Hinds-Levitt (HL) set:
distance power-law potentials, 92–93
double Z-score filter, THreading Onion Model 2 (THOM2) testing, 114–118
energy parameter optimization, 87–88
minimal models, 91–92
parameter-free optimization, 90–91
THOM1 potential, 94–100
Z-score filter testing, 111–112
HLA-DR1 pockets, protein-protein interactions, binding affinity evaluation, 440–442
Holm and Sander single decoys, native *vs.* decoy protein conformations, 472–474
Homologous proteins:
energy gap deletions, 107–108
gap penalty training, 103–107
HL test of, 116–118
protein folding statistical analysis, 28–29
tertiary protein structure, knowledge-based prediction, 214–218
identification, 215
local templates, 217–218
Homology modeling:
protein-protein interactions, binding site structure prediction, 407–408
sequence-structure-function paradigm, 147
sequence-structure-function prediction, 134, 155–165
atomic model reconstruction, 160–161
average model construction, 160
Fischer database results, 161–164
lattice chain, starting construction, 158
lattice folding optimization, 158–159
lattice model and force field, 157–158
Modeller tool, 164–165
Hooke's law, oligopeptide structure prediction, potential energy models, 289–291
Hydration shell, oligopeptide structure prediction:
computational studies, 302–312
solvation energy models, 292–296
Hydrophobic contacts:
native *vs.* decoy protein conformations, Park and Levitt decoys, 470–472
tertiary protein structure:
ab initio predictions, folding potential, 226–230
local templates, 218

Hydrophobic-polar model, protein folding
kinetics, 40
Hydrophobic residue-based superstructure,
polypeptide structure prediction, 360–
362
Hydrophobic zipper hypothesis, protein folding
statistical analysis, 25–26

IMPACT modeling program, native *vs.* decoy
protein conformations, 462–464
Interior dielectric constant, native *vs.* decoy
protein conformations, 481–482
Interior point algorithm, energy parameter
optimization, 89
Isothermal simulations, sequence-structure-
function prediction, 144–146
Iterative threading, sequence-structure-function
prediction, 151–152
genome-scale threading, 152–153

JPRED structures, tertiary protein simulations,
247
protocols, 248–250
JRF parameters, oligopeptide structure
prediction:
computational studies, 301–312
global minimization, 296–300
solvation energy models, 294–296

Kinetic partitioning mechanism (KPM), protein
folding mechanisms, 52–53
Knowledge-based prediction:
native *vs.* decoy conformations
applications, 483–484
approximate effective dielectric models:
distance-dependent dielectric
approximation, 480–481
screened Coulomb approximation,
478–480
CASP3 targets, 473–474
decoy data sets, 464–465
energy components, 474–478
force feld calculations, 462–464
Holm and Sander single decoys, 472–473
interior dielectric constant, 481–482
Park and Levitt decoys, 465–472
research background, 460–462
tertiary structure:
constraint methods, 201–210
ambiguous constraints, 205, 208–209
angle constraints, 202, 207–208
distance constraints, 202, 206–207
implementation, 204–205
predictions, derivation from, 203–204,
209–210
constraint refinement, 219
homology and structural templates,
214–218
identification, 215
local templates, 217–218
multiple templates, 215–217
protein modeling, 197–201
computational models, 197–198
geometrical representations, 198–199
scoring functions, 200–201
search algorithms, 199–200
research issues, 194–197
search space limitation, 210–214
fragment screening and enrichment,
212–213
local threading and fragment lists,
211–212
secondary structure modeling, 213–214
threading principle, 210–211
sequence-specific potentials, 218–219
Kronecker delta function, protein folding
kinetics, lattice models, 39

Lattice models:
disulfide bonds in protein folding, 58–62
protein folding kinetics, 6–8, 29–30, 38
basic principles, 38–41
computational models, 41–43
contact energies, 39
dense sequence space, 44–49
exhaustive enumeration, 41
Go model, 40–41
hydrophobic-polar model, 40
intermediates, 69–70
Monte Carlo simulation, 41–42
multiple histogram technique, 42
random bond model, 40
side chain models, 41
statistically derived pairwise potentials, 40
sequence-structure-function prediction:
ab initio prediction, 140–143
folding optimization, 158–159
force field parameters, 157–158
future issues, 180–182
lattice chain, starting construction, 158

Lattice models: (*Continued*)
tertiary protein structure, search algorithm, 199–200
Learning sets, energy parameter optimization, 87–88
Length cutoff, protein folding, coil-to-helix transition pathways, 374–378
Lennard-Jones potentials:
distance power-law potentials, 92–93
oligopeptide structure prediction, potential energy models, 290–291
pairwise potentials, 82–83
sequence-structure-function paradigm, *ab initio* prediction, 138–139
Levinthal paradox:
conformational space reduction, 43–44
protein folding kinetics, 36–38, 267–269
Ligand identification:
protein-protein interactions, binding affinity prediction, 410–411
sequence-structure-function prediction, low-resolution structures, 178–180
future improvements, 184
Linear programming:
energy parameter optimization, 88–89
pairwise potentials:
distance power-law potentials, 92–93
minimal models, 91–92
new energy profiles, 93–100
parameter-free models, 90–91
protein recognition, energy parameter calculations, 79–80
THOM1 potential, 94–100
threading potentials, 101–102
Linear regression, protein folding statistical analysis, 4–5
single-descriptor models, 16–19
Local alignments:
protein energy gaps, 108
double Z-score filter, 113–114
tertiary protein structure, search space limitation, local threading, 211–212
Local minimum energy conformations, oligopeptide structure prediction, 318–321
Local templates, tertiary protein structure, 217–218
Long helices, *ab initio* folding simulation, 257–258
Loop formation probability, protein folding:
kinetics, 56–58
proximity rule, 55
LOOPP program, protein recognition, gap penalties, THOM1 and THOM2 models, 106–107
Loop segments:
ab initio folding simulation, 257–258
tertiary protein structure, hierarchical search space limitation, 212–213
Low-resolution structures, sequence-structure-function prediction, ligand identification, 178–180
future improvements, 184

M. genitalium, sequence-structure-function prediction:
future issues, 183–186
genome-scale threading, 152–153
Major histocompatibility complex (MHC), protein-protein interactions:
binding affinity prediction, 414
computational studies, 440–442
decomposition techniques, 431
HLA-DR1 pockets, 440–442
minimization schemes, 438–440
modeling procedures, 431–438
binding site structure prediction, 407–408, 415–418
HLD-DR3 binding sites, 424–427
Many-body effects, pairwise interactions, 99–100
Master equation, coil-to-helix transition, time-evolution of quantities, 380–382
Met-enkephalin, oligopeptide structure prediction, computational studies, 304–312, 326–336
Metropolis criterion:
sequence-structure-function prediction, 145–146
tertiary protein structure, scoring functions, 200–201
Minimal models:
coil-to-helix transition:
pathway determination, 374–378
rate disconnectivity graph, 378–380
optimization, 91–92
protein folding, transition state algorithms, 393–397
tertiary protein structure:
geometrical representation, 198–199

search algorithm, 199–200
Minimum energy structure (MES), protein folding kinetics, dense sequence space, 46–49
Minimum-saddle-minimum triples, protein folding:
potential energy surfaces, 366–367
stationary points searching, 395
Misthreaded structures, THOM1 potential, 94–100
MODELLER technique, sequence-structure-function prediction:
comparative modeling, 134, 164–165
Fischer database results, 161–164
Moderate protein folding, mechanisms, 52–53
Molecular dynamics, oligopeptide structure prediction, solvation energy models, 292–296
Molecular mechanics models, oligopeptide structure prediction, potential energy models, 288–291
Monod-Wyman-Changeaux model, chaperonin-facilitated protein folding mechanism, 65–66
MONSSTER *ab initio* folding algorithm, sequence-structure-function prediction, 174–175
reduced/atomic model compatibility, 171–172
threading-based prediction, 153–155
Monte Carlo plus minimization (MCM) strategy:
oligopeptide structure prediction, local minimum energy conformations, 321
tertiary folding simulation:
PDB-derived and ideal structures, 238–241
predicted-derived secondary structure protocols, 248–250
Monte Carlo simulation:
oligopeptide structure prediction, solvation energy models, 292–296
protein folding kinetics, 6–8
lattice models, 38–41, 41–42
protein-protein interactions, binding affinity prediction, 412–414
sequence-structure-function prediction:
ab initio prediction, 139–143
sampling structure selection, 143–146
threading procedures, 136–137
Motif library, tertiary protein structure, search space limitation, local threading, 211–212
MSEED algorithm:
oligopeptide structure prediction:
global optimization, 299–300
solvation energy models, 293–296
protein-protein interactions:
binding site structure prediction:
modeling techniques, 435–436
solvation modeling, 437–438
global optimization, 420–421
Multiple-descriptor models, protein folding kinetic statistical analysis, 19–24
three-descriptor models, 22–24
two-descriptor models, 20–22
Multiple histogram, protein folding kinetics, lattice models, 42
Multiple sequence alignment (MSA) generation, sequence-structure-function prediction, first-pass threading, 148–149
Multiple templates, tertiary protein structure, knowledge-based prediction, 215–217
Murine binding site prediction, protein-protein interactions, 427–430
Myoglobin (1MBA), tertiary protein structure, distance constraints, 206–207
Myoglobin (1MBO), tertiary protein structure:
ambiguous constraints, 208–209
angle constraints, 207–208
structural template identification, 215

Native protein structures:
deletions, 107–108
disulfide bonds in protein folding:
bovine pancreatic trypsin inhibitor (BPTI), 55
modeling techniques, 58–62
energy parameter optimization, 86–87
gap energy optimization, 101–107
THOM1 potential, decoy structures, 97–100
THOM2 potential:
HL test, 118
self-recognition of folds, 120
vs. decoy conformations:
applications, 483–484
approximate effective dielectric models:
distance-dependent dielectric approximation, 480–481

Native protein structures: (*Continued*)
screened Coulomb approximation, 478–480
CASP3 targets, 473–474
decoy data sets, 464–465
energy components, 474–478
force feld calculations, 462–464
Holm and Sander single decoys, 472–473
interior dielectric constant, 481–482
Park and Levitt decoys, 465–472
research background, 460–462
Nearly bimodal distrbution (NBD), two-state protein folding, 49–51
Needleman-Wunsch global alignment program, sequence-structure-function prediction, first-pass threading, 149
Negative energies, Z-score filter testing, 109–112
Neural networks:
folding rates of proteins, 9–11
protein folding statistical analysis, 4–5
single-descriptor models, 16–19
Newton-Raphson method, protein folding:
searching strategies, 365–366
stationary points searching, 392–393
Nonbonded energy terms, oligopeptide structure prediction, potential energy models, 290–291
Non-native intermediates, disulfide bonds in protein folding:
bovine pancreatic trypsin inhibitor (BPTI), 54–55
modeling techniques, 58–62
NPSOL program:
protein-protein interactions:
binding site structure prediction, 435–436
global optimization, 438–440
global optimization, 420–421
sparse restraints, structural refinement, αBB algorithm, 345–347
Nuclear magnetic resonance (NMR):
sparse restraints:
computational study, 347–353
structural refinement, 338–359
tertiary protein structure, constraint analysis, 201–210
Nuclear Overhauser effect, sparse restraints, structural refinement:
computational study, 348–353
energy modeling, 339–341

Oligopeptide structural prediction, deterministic global optimization:
computational models, 301–312
free energy computational studies, 322–336
free energy modeling, 312–314
global optimization framework, 296–300
harmonic approximation, 314–318
local minimum energy conformation, 318–322
potential energy models, 288–291
solvation energy models, 291–296
1aj3 protein, *ab initio* folding simulation, 256
1am3 protein, *ab initio* folding simulation, 256–257
1bg8-chain A protein, *ab initio* simulations, 259–260
1mzm protein, *ab initio* folding simulation, 257
OONS parameter set, oligopeptide structure prediction:
computational studies, 301–312
solvation energy models, 293–296
Open reading frames (ORFs), sequence-structure-function paradigm, 132
genome-scale threading, 153
OPLS-AA force field, native *vs.* decoy conformations:
applications, 483–484
approximate effective dielectric models:
distance-dependent dielectric approximation, 480–481
screened Coulomb approximation, 478–480
CASP3 targets, 473–474
decoy data sets, 464–465
energy components, 474–478
force feld calculations, 462–464
Holm and Sander single decoys, 472–473
interior dielectric constant, 481–482
Park and Levitt decoys, 465–472
research background, 460–462
Optimization cycle. *See also* Deterministic global optimization
tertiary protein structure, *ab initio* folding prediction, potential energy function, 231–237

PACK program:
oligopeptide structure prediction, global optimization, 299–300

protein-protein interactions:
binding site structure prediction, 435–436
global optimization, 438–440
global optimization, 420–421
sparse restraints, structural refinement:
αBB algorithm, 345–347
computational analysis, 349–353
Pair energies, THOM2 recognition *vs.*, 120–126
Pair interactions:
sequence-structure-function prediction, threading procedures, 135–136
statistics of, 98–100
Pairwise potentials:
distance power-law potentials, 92–93
many-body effects, 99–100
minimal models, 91–92
new energy profiles, 93–100
parameter-free models, 90–91
protein folding kinetics, 40
THreading Onion Model (THOM), 81–83
threading potentials, protein recognition, 79–80
Parallelization, deterministic global optimization, 363
Parameter-free models, optimization, 90–91
Park and Levitt decoys:
native *vs.* decoy protein conformations, 466–472
energy components, 476–478
interior dielectric constant, 481–482
"Partially thawed" approximation, sequence-structure-function prediction, first-pass threading, 148–149
Pathway determination, protein folding:
coil-to-helix transition, 373–378
potential energy surface, 400–402
Pearson linear correlation coefficient, protein folding statistical analysis, 3–4
lattice models, 7–8
single-descriptor models, 16–19
Penalty functions, tertiary protein structure, constraint implementation, 204–205
PhD structures, tertiary protein simulations, 247
Phenomenological models:
folding rates of proteins, 10–11
protein folding kinetics, 37–38
Pocket definition, protein-protein interactions, binding site structure prediction, 416–418
global optimization, 420–421
HLA-DR1 pockets, 440–442
modeling techniques, 431–435
Poisson-Boltzmann equation:
native *vs.* decoy protein conformations:
SGB calculations, 464
solvation effects, 461–462
oligopeptide structure prediction, solvation energy models, 292–296
protein-protein interactions, binding site structure prediction, 443–444
Polypeptide structural prediction, deterministic global optimization:
computational models, 301–312
free energy computational studies, 322–336
free energy modeling, 312–314
future issues, 360–362
global optimization framework, 296–300
harmonic approximation, 314–318
local minimum energy conformation, 318–322
potential energy models, 288–291
solvation energy models, 291–296
Potential energy global minimum (PEGM), oligopeptide structure prediction, computational studies, 322–336
Potential energy surface:
oligopeptide structure prediction, 288–291
protein folding mechanism:
equilibrium probabilities, 398–399
free energy calculation, 398
global optimization, 365–367
pathway definition, 400–402
rate disconnectivity graph, 402–403
time-dependent probabilities, 399–400
transition rates, 399
vibrational frequencies calculation, 397–398
tertiary protein structure, *ab initio* folding potential predictions, 230–237
Potential of mean force (pmf), tertiary protein structure:
ab initio folding potential predictions, potential energy function, 230–237
ab initio predictions, folding potential, 226–230
POU proteins, THOM2 model *vs.* pair energies, 120–126

Predicted-derived secondary structure, *ab initio* predictions, folding potential, tertiary protein structure simulations, 246–260
ranking procedures, 250–260
simulation protocols, 248–250
techniques, 247
Prediction techniques:
knowledge-based prediction, tertiary structure:
constraint methods, 201–210
ambiguous constraints, 205, 208–209
angle constraints, 202, 207–208
distance constraints, 202, 206–207
implementation, 204–205
predictions, derivation from, 203–204, 209–210
constraint refinement, 219
homology and structural templates, 214–218
protein modeling, 197–201
research issues, 194–197
search space limitation, 210–214
sequence-specific potentials, 218–219
native *vs.* decoy conformations, 460–462
applications, 483–484
approximate effective dielectric models:
distance-dependent dielectric approximation, 480–481
screened Coulomb approximation, 478–480
CASP3 targets, 473–474
decoy data sets, 464–465
energy components, 474–478
force feld calculations, 462–464
Holm and Sander single decoys, 472–473
interior dielectric constant, 481–482
Park and Levitt decoys, 465–472
research background, 460–462
polypeptide structure prediction:
computational models, 301–312
free energy computational studies, 322–336
free energy modeling, 312–314
global optimization framework, 296–300
harmonic approximation, 314–318
local minimum energy conformation, 318–322
potential energy models, 288–291
solvation energy models, 291–296
sequence-structure-function:
ab initio prediction, 138–143
biochemical function, 174–175
methodology, 165–167
test protein results, 167–169
biochemical function, transition to, 172–178
ab initio folding, 174–175
active site identification, 172–174
future issues, 183–184
threaded structures, 175–178
comparative modeling methods, 134, 155–165
atomic model reconstruction, 160–161
average model construction, 160
Fischer database results, 161–164
lattice chain, starting construction, 158
lattice folding optimization, 158–159
lattice model and force field, 157–158
Modeller tool, 164–165
experimental data and, 182–183
future improvements, 180–186
low-resolution structures, ligand identification, 178–180
future improvements, 184
reduced/atomic model compatibility:
atomic reconstruction, 170–171
structural refinement feasibility, 171–172
structural reproducibility, 169–170
research background, 132–134
sampling techniques, 143–146
threading procedures, 134–138
first-pass threading, 148–149
Fischer database applications, 149–151
genome-scale iterative threading, 152–153
iterative threading, 151–152
orientation-dependent pair potential, PROSPECTOR extension, 153
tertiary contacts, 153–155
unified folding method, 146–148
tertiary protein structure, constraint prediction, 209–210
prePACK program, protein-protein interactions, binding site structure prediction, 420–421
PRODOCK software, protein-protein interactions, binding affinity prediction, 412–414

"Progress variables," disulfide bonds in protein folding, 53–55
PROSITE patterning, tertiary protein structure, structural template identification, 215
PROSPECTORIEN1-3, sequence-structure-function prediction:
  biochemical function prediction, 176–178
  threading-based prediction, 153–155
PROSPECTOR/PROSPECTOR2 algorithms, sequence-structure-function prediction:
  biochemical function prediction, 175–178
  first-pass threading, 148–149
  Fischer database:
    applications, 149–152
    results, 161–164
  future issues, 180–186
  orientation-dependent pair potential, 153
  threading procedures, 137–138
  unified folding method, 147–148
PROSPECTORQUASI1-3 algorithm, sequence-structure-function prediction, genome-scale threading, 152–155
Protein Data Bank (PDB):
  native *vs.* decoy protein conformations:
    decoy data sets, 465
    force field calculations, 462–464
  sparse restraints, structural refinement, computational analysis, 349–353
  tertiary protein structure:
    ab initio predictions, folding potential, 224–226
      derived and ideal secondary structures, 237–246
      potential energy function, 233–237
      simulation errors, 226–230
    constraint analysis, 203–205
    knowledge-based prediction technique, 195–196
    scoring functions, 200–201
    structural template identification, 215
    threading protocol, search space limitation, 210–211
Protein disulfide isomerase (PDI), protein folding kinetics, 58
Protein folding mechanisms:
  chaperonin mechanisms, 63–69
    GroEL unfolding, 66–68
    stretching, unfolding by, 68–69
  deterministic global optimization:
    αBB algorithm, 367–369
  coil-to-helix transitions, 369–390
    pathways, 373–378
    rate disconnectivity graphs, 378–380
    reaction coordinates, 383–386
    solvated tetra-alanine, 386–390
    time evolution of quantities, 380–382
    transition rates and master equation, 371–373
    unsolvated tetra-alanine, 370–371
  framework and implementations, 390–403
    local stationary point search methods, 392–393
    minima and transition rate search, 393–397
    potential energy surface analysis, 397–403
  future research, 403–405
  potential energy surface, 366–367
  secondary structure formation, 364
  stationary points, 365–366
disulfide bonds:
  barnase engineering, 62–63
  BPTI refolding, 53–55
  modeling techniques, 58–62
  proximity rule, 55–58
moderate folders, topologcal frustration, kineetic partitioning mechanism, 52–53
native *vs.* decoy conformations:
  applications, 483–484
  approximate effective dielectric models:
    distance-dependent dielectric approximation, 480–481
    screened Coulomb approximation, 478–480
  CASP3 targets, 473–474
  decoy data sets, 464–465
  energy components, 474–478
  force feld calculations, 462–464
  Holm and Sander single decoys, 472–473
  interior dielectric constant, 481–482
  Park and Levitt decoys, 465–472
  research background, 460–462
proximity rules, 56–58
statistical analysis:
  folding rates, 9–26
    database, 11–16
    multiple-descriptor models, 19–24
    physical observations, 24–26
    review, 9–11
    single-descriptor models, 16–19

Protein folding mechanisms: (*Continued*)
homologous proteins, 28–29
lattice models, 6–8
protein and lattice model studies, 29–30
research background, 2–3
techniques, 3–6
unfolding rates, 26–28
two-state folders, 49–51
Protein-protein interactions, deterministic global optimization:
binding affinity prediction, 408–414, 430–442
computational studies, 440–442
decomposition techniques, 431
minimization scheme, 438–440
modeling, 431–438
binding site structure prediction, 407–408, 414–430
class II HLA molecule structures, 415
computational studies, 421–430
modeling techniques, 418–420
prediction techniques, 416–418
programming tools, 420–421
future research issues, 442–444
research background, 404–414
Protein recognition, sequence-to-structure fitness:
deletion energy optimization, 107–108
energy parameter optimization, 85–89
learning and control sets, 87–88
linear programming protocol, 88–89
folds excluded from training, 118–120
gap energy optimization, 100–107
HL model test, 114–118
LOOPP program, 126–127
pair and profile energy evaluation, 89–100
distance power-law potentials, evaluation, 92–93
"minimal" models, 91–92
parameter-free models, 90–91
THOM models capacity, 93–100
pairwise models, 81–83
profile models, 83–85
research background, 78–81
self-recognition of folds, 118–120
statistical significance of results, 108–114
double Z-score filter, 112–114
Z-score filter, 108–112
THOM2 *vs.* pair energies, 120–126
Proteome, interpretation techniques, 132
Proximity rule:
disulfide bonds in protein folding:
bovine pancreatic trypsin inhibitor (BPTI), 55
folding kinetics, 56–58
loop formation probability, 55
protein folding kinetics, 38
Pseudo-native structures, gap energy optimization, 101–107
PSI-BLAST sequence profiling, sequence-structure-function prediction:
Fischer database applications, 151
threading procedures, 137
unified folding method, 146–148
PSIPRED techniques:
ab initio simulations, final rankings, 250–260
predicted-derived secondary structure simulations, tertiary protein structures, 246–247

Quantitative structure-property relationships (QSPRs), protein folding statistical analysis, lattice models, 8

Random alignment, energy gaps, Z-score filter, 109–112
Random bond model, protein folding kinetics, 40
dense sequence space, 45–49
Randomized data, protein folding statistical analysis, two-descriptor models, 21–22
Rate disconnectivity graph, protein folding:
coil-to-helix transition, 378–380
potential energy surface, 402–403
Reaction coordinates, protein folding:
coil-to-helix transition, 383–386
potential energy surface, 400–402
proximity rule, 55
Recursion techniques, sparse restraints, structural refinement, 343–345
Reduced/atomic model compatibility, sequence-structure-function prediction:
atomic reconstruction, 170–171
structural refinement feasibility, 171–172
structural reproducibility, 169–170
Reduced Radius Independent Gaussian Sphere (RRIGS) approximation:
oligopeptide structure prediction:
computational studies, 301–312
energy-direct approach (EDA), 331–336

global optimization, 299–300
solvation energy models, 296–297
protein folding dynamics, coil-to-helix transitions, 369–370
coil-to-helix transitions, 387–390
protein-protein interactions, binding site structure prediction, 435–436
global optimization, 438–440
solvation modeling, 437–438
Replica exchange Monte Carlo (REMC), sequence-structure-function prediction, 146
restrained lattice folding, 158–159
Restrained lattice folding, sequence-structure-function prediction, 158–159
Rice-Ramsperger-Kassel-Marcus (RRKM) theory, protein folding:
coil-to-helix transition, 371–373
potential energy surfaces, 367
transition rate cutoff, 399
Rigid models, protein-protein interactions, binding affinity prediction, 410–411
Root-mean-square-deviation (RMSD):
native *vs.* decoy protein conformations:
CASP3 targets, 474–475
decoy data sets, 464–465
dielectric models, screened Coulomb approximation, 478–480
energy components, 475–478
Holm and Sander single decoys, 473–474
interior dielectric constant, 481–482
Park and Levitt decoys, 466–472
protein-protein interactions, binding site structure prediction, cRMSD, 423–430
sequence-structure-function paradigm:
ab initio prediction, 138–139
folding, 167–169
Modeller comparisons, 164–165
reduced/atomic model compatibility, 169–172
threading-based prediction, tertiary contracts, 153–155
sparse restraints, structural refinement:
computational analysis, 350–353
DYANA protocol, 353–356
tertiary protein structure, *ab initio* folding prediction, 225–226
PDB-derived and ideal structures, 238–241
potential energy function, 231–237
ROSETTA method, sequence-structure-function prediction:
ab initio prediction, 141–143
future issues, 180–182
Rotation matrices, protein-protein interactions, binding site structure prediction, 432–435
RUBISCO protein, chaperonin-facilitated protein folding mechanism, stretching-induced unfolding, 68–69

Sampling procedures, sequence-structure-function prediction, 143–146
SCKS global minimum structure, computational studies, 308–312
SCOP database, sequence-structure-function prediction, 172
Scoring functions:
native protein detection, research background, 460–462
native *vs.* decoy protein conformations, Park and Levitt decoys, 468–472
protein-protein interactions, binding affinity prediction, 408–414
tertiary protein structure, 200–201
constraint implementation, 204–205
Search algorithms, tertiary protein structure, 199–200
Search space limitation, tertiary protein structure, knowledge-based prediction, 210–214
fragment screening and enrichment, 212–213
local threading and fragment lists, 211–212
secondary structure modeling, 213–214
threading principle, 210–211
Secondary protein structure:
fragment-based modeling, tertiary protein structure, 213–214
protein folding mechanism, global optimization, 364
tertiary folding simulation:
PDB-derived and ideal structures:
methodology, 237–241
physical model, 237
size-dependent potential, 241–245
terminal loop definition and truncation, 245
predicted-derived secondary structure, 246–260
ab initio predictions, 250–260

Secondary protein structure: (*Continued*)
simulation protocols, 248–250
techniques, 247
SEGMOD program, native *vs.* decoy protein conformations, decoy data sets, 464–465
Self-recognition, sequence-to-structure compatibility:
folds excluded from training, 118–120
HL test, 114–118
Sequence alignment:
sequence-structure-function paradigm, 146–147
tertiary protein structure, constraint analysis, 203–204
Sequence-specific potentials, tertiary protein structure, 218–219
Sequence-structure-function paradigm:
prediction techniques:
ab initio prediction, 138–143
biochemical function, 174–175
methodology, 165–167
test protein results, 167–169
biochemical function, transition to, 172–178
ab initio folding, 174–175
active site identification, 172–174
future issues, 183–184
threaded structures, 175–178
comparative modeling methods, 134, 155–165
atomic model reconstruction, 160–161
average model construction, 160
Fischer database results, 161–164
lattice chain, starting construction, 158
lattice folding optimization, 158–159
lattice model and force field, 157–158
Modeller tool, 164–165
experimental data and, 182–183
future improvements, 180–186
low-resolution structures, ligand identification, 178–180
future improvements, 184
reduced/atomic model compatability:
atomic reconstruction, 170–171
structural refinement feasibility, 171–172
structural reproducibility, 169–170
research background, 132–134
sampling techniques, 143–146
threading procedures, 134–138
first-pass threading, 148–149
Fischer database applications, 149–151
genome-scale iterative threading, 152–153
iterative threading, 151–152
orientation-dependent pair potential, PROSPECTOR extension, 153
tertiary contacts, 153–155
unified folding method, 146–148
protein recognition:
deletion energy optimization, 107–108
energy parameter optimization, 85–89
learning and control sets, 87–88
linear programming protocol, 88–89
gap energy optimization, 100–107
HL model test, 114–118
LOOPP program, 126–127
pair and profile energy evaluation, 89–100
distance power-law potentials, evaluation, 92–93
"minimal" models, 91–92
parameter-free models, 90–91
THOM models capacity, 93–100
pairwise models, 81–83
profile models, 83–85
research background, 78–81
self-recognition of folds, 118–120
statistical significance of results, 108–114
double Z-score filter, 112–114
Z-score filter, 108–112
THOM2 *vs.* pair energies, 120–126
Shifted Lennard-Jones potential (SLJ):
distance power-law potentials, 92–93
protein recognition, 82–83
Side chain conformation:
native *vs.* decoy protein conformations, 469–472
tertiary protein structure, 198–199
Side chain-only (SICHO) models, sequence-structure-function prediction:
atomic model reconstruction, 160–161
future issues, 182–186
lattice model force field, 157–158
reduced/atomic model compatibility, 169–172
Side chains, protein folding kinetics, lattice models, 41
Simulated annealing, protein-protein interactions, binding affinity prediction, 412–414

Simulation errors, tertiary protein structure, *ab initio* folding potential predictions, 227–230
Single-descriptor models, protein folding statistical analysis, 16–19
  linear correlations, 16
  neural network predictions, 16–19
Single-domain proteins:
  folding kinetics, 36–38
  native *vs.* decoy protein conformations, 472–474
Single-step potentials, protein recognition, THreading Onion Model (THOM), 81–83
Site mapping, protein-protein interactions, binding affinity prediction, 413–414
Size-dependent potential energy function, tertiary protein structure, *ab initio* predictions, folding potential, 226–237
  error identification, 226–230
  improvements in, 230–233
  PDB-derived secondary structure, 241–245
  results, 233–237
Smith-Waterman algorithm, THOM2 model *vs.* pair energies, 120–126
Solvation energy models:
  native *vs.* decoy protein conformations:
    Coulomb approximation, dielectric models, 478–480
    research background, 460–462
  oligopeptide structure prediction, 291–296
    computational studies, 301–312
  protein-protein interactions, binding site structure prediction, 416–418, 437–438
Sparse restraints, structural refinement, 336–359
  algorithmic steps, 345–347
  computational study, 347–353
  DYANA *vs.* torsion angle dynamics (TAD), 353–356
  energy modeling, 339–341
  global optimization, 341
  torsion angle dynamics (TAD), 341–345, 356–359
Square well potential (SWP), capacity comparisons of new models, 93–100
Stationary points, protein folding mechanism:
  clustering algorithm, 396–397
  eigenmode-following search, 392–393
  fill algorithm, 395–396
  Grid search, 394
  research framework, 391–393
  searching strategies, 365–366
  SUMSL algorithm, 393
  triples search, 395
  unsolvated tetra-alanine, 370–371
  uphill search, 394–395
Statistical analysis, protein folding kinetics:
  folding rates, 9–26
    database, 11–16
    multiple-descriptor models, 19–24
    physical observations, 24–26
    review, 9–11
    single-descriptor models, 16–19
  homologous proteins, 28–29
  lattice models, 6–8
  protein and lattice model studies, 29–30
  research background, 2–3
  techniques, 3–6
  unfolding rates, 26–28
Structure prediction techniques:
  polypeptides:
    computational models, 301–312
    free energy computational studies, 322–336
    free energy modeling, 312–314
    global optimization framework, 296–300
    harmonic approximation, 314–318
    local minimum energy conformation, 318–322
    potential energy models, 288–291
    solvation energy models, 291–296
  sequence-structure-function paradigm, 147–148
  tertiary protein structure, knowledge-based prediction, 214–218
    identification, 215
    local templates, 217–218
    multiple templates, 215–217
Substrate protein (SP), chaperonin-facilitated protein folding mechanism, 64–66
  stretching-induced unfolding, 68–69
SUMSL algorithm, protein folding, stationary points searching, 393
Surface formulation of generalized Born model (SGB), native *vs.* decoy conformations:
  applications, 483–484
  approximate effective dielectric models:
    distance-dependent dielectric approximation, 480–481

Surface formulation of generalized Born model (SGB), native *vs.* decoy conformations: (*Continued*)
screened Coulomb approximation, 478–480
CASP3 targets, 473–474
decoy data sets, 464–465
energy components, 474–478
force feld calculations, 462–464
Holm and Sander single decoys, 472–473
interior dielectric constant, 481–482
Park and Levitt decoys, 465–472
research background, 460–462

Template competition, tertiary protein structure, 216
Tendamistat (3AIT), tertiary protein structure, distance constraints, 206–207
TE potential:
capacity comparisons of new models, 93–100
distance power-law potentials, 92–93
energy parameter optimization, 87–88
minimal models, 91–92
native alignments in, 98–100
parameter-free optimization, 90–91
THOM2 model *vs.* pair energies, 125–126
Terminal loops, tertiary folding simulation, secondary structure definition and truncation, 245
Tertiary contacts, sequence-structure-function prediction:
restrained lattice folding, 158–159
threading-based prediction, 153–155
Tertiary protein structure:
ab initio predictions, folding potential:
PDB-derived secondary structure simulations, 238–246
methodology, 238–241
size-dependent potential, 241–245
terminal loop definition and truncation, 245
three-dimensional topologies, 245–246
predicted-derived secondary structure simulations, 246–260
research background, 224–226
size-dependent potential energy function, 226–237
error identification, 226–230
improvements in, 230–233
results, 233–237
knowledge-based prediction:
constraint methods, 201–210
ambiguous constraints, 205, 208–209
angle constraints, 202, 207–208
distance constraints, 202, 206–207
implementation, 204–205
predictions, derivation from, 203–204, 209–210
constraint refinement, 219
homology and structural templates, 214–218
identification, 215
local templates, 217–218
multiple templates, 215–217
protein modeling, 197–201
computational models, 197–198
geometrical representations, 198–199
scoring functions, 200–201
search algorithms, 199–200
research issues, 194–197
search space limitation, 210–214
fragment screening and enrichment, 212–213
local threading and fragment lists, 211–212
secondary structure modeling, 213–214
threading principle, 210–211
sequence-specific potentials, 218–219
Tetra-alanine:
coil-to-helix transition, 370–371
pathway determination, 374–378
rate disconnectivity graph, 378–380
protein folding, coil-to-helix transition, solvated model, 386–390
Thermodynamic hypothesis:
protein folding mechanisms, 268–269
sequence-structure-function prediction, 144–146
tertiary protein structure, computational modeling, 197–198
THreading Onion Model 1 (THOM1):
capacity comparisons, 93–100
dissection of potential, 94–100
energy parameter optimization, linear programming, 88–89
energy profiles, 83–85
gap energy training, 104–107
minimal models, 91–92
pairwise models, 81–83
parameter-free models, 90–91

THreading Onion Model 2 (THOM2):
capacity comparisons, 93–100, 98–100
energy profiles, 83–85
gap energy training, 104–107
minimal models, 91–92
pairwise models, 81–83
testing of, 114–126
HL test, 114–118
protein recognition excluded from training, 118–120
vs. pair energies, 120–126
Threading protocols:
comparison of potentials, 91–92
linear programming (LP) protocol, 101–102
protein recognition, 78–79
sequence-structure-function prediction, 134–138, 147
biochemical function prediction, 175–178
first-pass threading, 148–149
Fischer database applications, 149–151
genome-scale iterative threading, 152–153
iterative threading, 151–152
orientation-dependent pair potential, PROSPECTOR extension, 153
tertiary contacts, 153–155
tertiary protein structure, search space limitation, knowledge-based prediction, 210–211
local threading and fragment lists, 211–212
THOM2 model, HL test, 114–118
Three-descriptor models, protein folding kinetic statistical analysis, 22–24
Time-dependent probabilities, protein folding, potential energy surfaces, 399–400
Time-evolution of quantities, protein folding, coil-to-helix transition, 380–382
Topological frustration, protein folding mechanisms, 52–53
Torsional energy expressioin, oligopeptide structure prediction, potential energy models, 289–291
Torsion angle dynamics (TAD), sparse restraints, structural refinement, 338–359
algorithmic steps, 345–347
global optimization, 356–359
molecular dynamics, 341–345
vs. DYANA protocol, 353–356
Total energy profile, protein recognition, 83–85
Transition rate cutoff, protein folding:
coil-to-helix transition pathways, 374–378
potential energy surface calculation, 399
Transition states, protein foldinig, minima and first-order algorithms, 393–397
Translation vector, protein-protein interactions, binding site structure prediction, 432–435
Triose phosphate isomerase (TIM), sequence-structure-function prediction, 172
Triples algorithm, protein folding:
potential energy surfaces, 366–367
stationary points searching, 395
Twice continuously differentiable NLPs, deterministic global optimization, 269–279
αBB algorithm, 276–279
convex lower bounding, 275
feasible region convexification, 274–275
nonconvex terms, underestimation, 272–274
special structure, underestimation, 270–272
variable bound updates, 275–276
Two-descriptor models, protein folding kinetic statistical analysis, 20–22
Two-state protein folding:
kinetics, 37–38
mechanisms, 49–51

Underestimating problems, deterministic global optimization, twice continuously differentiable NLPs, αBB algorithm, 277–279
Unfolding rates:
chaperonin-facilitated protein folding mechanism, 66–67
stretching techniques, 68–69
protein folding statistical analysis, 26–28
Unified folding method, sequence-structure-function prediction, 146–148
Univariate concave functions, deterministic global optimization, twice continuously differentiable NLPs, 270–272
Uphill search algorithm, protein folding, 394–395
Upper bound check (UBC), oligopeptide structure prediction:
global optimization, 299–300
local minimum energy conformations, 320–321
Urey-Bradley distance, potential energy surface, 404–405

Van der Waals interactions:
native *vs.* decoy protein conformations, 476–478
oligopeptide structure prediction, potential energy models, 290–291
Variable bound updates, deterministic global optimization, twice continuously differentiable NLPs, 275–276
αBB algorithm, 276–279
Vibrational frequencies, protein folding mechanism, potential energy surface, 397–398
Zero-parameter models, optimization, 90–91
Zimmerman codes, oligopeptide structure prediction, computational studies, 325–336
Z-score filter:
energy gap alignments, 108–112
native *vs.* decoy protein conformations, Park and Levitt decoys, 468–472
THOM2 model:
HL test, 114–118
vs. pair energies, 120–126