Journal of Mathematical Chemistry 14 (1993) 35–45

Protein secondary structure conformations and associated hydrophobicity scales

Davor Juretić [1]

Natural Sciences and Arts Department, University of Split, N. Tesle 12, 58000 Split, Croatia

Nenad Trinajstić and Bono Lučić

Rugjer Bošković Institute, 41000 Zagreb, Croatia

We have developed conformational preference functions and a hierarchy of algorithms that can evaluate the success of each hydrophobicity scale in predicting protein secondary conformation. The results of such evaluation are shown for fiftyfive different scales with respect to their ability to predict α-helix, β-sheet and coil structure in three testing sets of proteins: five integral membrane proteins, twelve α-class and sixteen β-class soluble proteins. Our scale of conformational parameters is the best predictor of secondary structure segments in membrane proteins and α-class proteins. The success rate and correlation coefficient for α-helix conformation in membrane proteins are 76% and 0.46 respectively, which is superior to the performance measures attained with other prediction schemes. Evaluation of solution hydrophobicity scales, often used to predict transmembrane segments in membrane proteins, indicated absence of correlation in prediction of helix segments and experimental results for the conformation of membrane proteins. Such scales have better performance (correlation coefficient around 0.30) in predicting sheet conformation in the β-class proteins.

1. Introduction

Hydrophobic energies or solvent effects are a major contributor to the energetics of protein folding [1]. Many different assignments of hydrophobicity values to amino acids (hydrophobicity scales) have been proposed in the literature [2]. Some of them have been used in attempts to improve accuracy of secondary structure prediction algorithms [3–6] or to identify transmembrane segments of membrane proteins [7–10]. It is unlikely that one scale will be the best for all protein classes and for all secondary structure conformations [11,12], but a systematic evaluation of scales has not been done before.

In this report we present the results of such an evaluation for fiftyfive different scales of conformational parameters with respect to their ability to predict α-helix, β-sheet and coil structure, not only in soluble proteins of α and β class, but also in

[1] To whom correspondence should be addressed.

several integral membrane proteins. One of our own scales is the best for predicting the structure of integral membrane proteins and soluble α-class proteins.

2. Methods

We use the method of conformational preference functions for predicting secondary structure conformation [13,14]. It is based on a modified sliding window procedure [15,16], that for each amino acid from the protein data base (same 100 soluble proteins of known structure that were used in ref. [14]) collects the information about its type, its secondary conformation and its "hydrophobic environment". The environment of a residue n in the sequence is defined as the average of a selected property over residues $n-4$ to $n+4$, excluding residue n. The frequency distributions over different environments for each amino acid type and in each secondary structure is approximately normal in accord with the central limit theorem [17]. A set of Gaussian functions that replaces all histograms over environments is used to construct preference functions [13].

In a protein of unknown structure, sequence dependent preferences for different conformations are compared for each residue and conformation is assigned to the highest preference. In a protein (or protein list) of known structure the correlation coefficient C_s [18] is used to report prediction accuracy for each secondary structure type s. The success rate, Q_3, which is the percentage of correctly predicted residues is also reported. The percentage of correctly predicted residues of structure type s: Q_{3s} for α-helix, β-sheet and coil conformation is reported too.

All programs are written in Fortran and are available from the first author. Three testing sets of proteins: five integral membrane proteins, twelve α-class proteins and sixteen β-class proteins are listed below. The structures and the sequences of α-class and β-class proteins known with 3 Å or better resolution were obtained from the Brookhaven Protein Data Bank (PDB) [19], while references used for the primary and secondary structure of membrane proteins are enclosed.

Integral membrane proteins: bacteriorhodopsin [20], bacterial photosynthetic reaction center L subunit [21,22], bacterial photosynthetic reaction center M subunit [21,22], lactose permease [23], and rhodopsin [24].

α-class soluble proteins (PDB file name): 155c, 156b, 1ccr, 1eco, 1mbd, 2cpp, 2lh7, 2lhb, 3cpv, 3icb, 451c, 1hho.

β-class soluble proteins (PDB file name): 1gcr, 1hip, 2stv, 3cna, 1acx, 1fbj, 1gpl, 1hmg, 2act, 2alp, 2est, 2rhe, 2sbt, 2sga, 2tbv, 4sbv.

3. Results and discussion

The main results are presented in table 1. Scales of chemical, physical or statistical properties of amino acids are listed there and in the appendix, in the order of

Table 1
Performance parameters for membrane proteins, α-class soluble proteins and β-class soluble proteins.

Scale	Membrane					α-class					β-class				
	Q_3	C_α	C_c	$Q_{3\alpha}$	Q_{3c}	Q_3	C_α	C_c	$Q_{3\alpha}$	Q_{3c}	Q_3	C_β	C_c	$Q_{3\beta}$	Q_{3c}
JTL	67	0.46	0.50	76	56	67	0.39	0.39	68	66	51	0.23	0.32	33	67
SCHERA	67	0.43	0.46	77	54	66	0.39	0.38	69	64	49	0.22	0.31	33	62
FASTUR	65	0.43	0.46	75	52	66	0.36	0.34	73	56	51	0.22	0.31	34	67
ROBSON	62	0.40	0.44	70	50	64	0.36	0.35	72	53	51	0.23	0.32	41	61
MIJER	51	0.33	0.52	51	53	54	0.29	0.35	43	70	56	0.26	0.32	53	67
ROSE1	61	0.31	0.42	68	52	59	0.28	0.30	57	63	50	0.23	0.33	36	65
KARPLU	62	0.30	0.30	66	56	62	0.30	0.30	66	56	52	0.25	0.32	38	66
RICH1	58	0.30	0.38	63	51	66	0.36	0.37	71	56	51	0.28	0.36	41	61
CHOU3	56	0.30	0.36	61	48	65	0.34	0.33	76	48	53	0.28	0.33	48	61
WERSC	49	0.30	0.53	47	53	52	0.27	0.31	41	68	57	0.27	0.32	52	68
RICH2	55	0.29	0.36	62	45	66	0.32	0.33	79	45	53	0.25	0.33	50	59
BULDG	52	0.29	0.51	52	51	56	0.31	0.35	47	72	55	0.25	0.32	48	68
DEBER	44	0.27	0.52	41	50	53	0.27	0.35	44	70	57	0.28	0.32	53	68
ABODR	39	0.26	0.44	35	25	50	0.26	0.35	40	67	58	0.30	0.32	64	62
GUYM	44	0.26	0.51	39	53	50	0.27	0.32	39	67	57	0.28	0.32	58	65
CHART	56	0.24	0.37	61	51	60	0.28	0.31	62	59	50	0.22	0.32	35	64
FASALF	50	0.24	0.36	54	46	66	0.35	0.34	78	46	53	0.27	0.33	50	58
HELL1	45	0.23	0.49	44	49	54	0.29	0.36	46	67	57	0.29	0.33	54	67
CHOU2	46	0.22	0.48	41	55	55	0.26	0.31	48	64	55	0.23	0.32	45	67
NEIL	55	0.21	0.34	56	55	63	0.32	0.34	67	61	42	0.17	0.23	27	53
COHEN	47	0.19	0.31	45	52	57	0.27	0.27	58	54	56	0.27	0.36	47	70
GIBRAT	37	0.18	0.51	29	52	51	0.26	0.35	41	68	58	0.30	0.32	59	64
KRIGK	38	0.18	0.50	28	54	50	0.26	0.35	38	71	58	0.30	0.33	62	63
CHOU4	47	0.17	0.34	48	46	62	0.33	0.34	71	52	54	0.29	0.34	50	60
PRIFT	34	0.16	0.51	24	50	49	0.26	0.36	35	75	58	0.29	0.32	51	73
SWEET	36	0.16	0.49	28	50	48	0.24	0.33	37	68	59	0.31	0.33	59	68
CHOU1	44	0.14	0.35	36	59	62	0.36	0.35	64	62	52	0.25	0.31	38	63
JONES	43	0.14	0.38	39	51	59	0.29	0.33	59	59	56	0.30	0.36	47	69
CHOTH1	48	0.11	0.32	51	45	59	0.27	0.27	66	49	49	0.21	0.34	40	57
PONNU	32	0.10	0.50	19	55	49	0.25	0.35	36	71	59	0.30	0.34	59	67
GUY	34	0.10	0.47	22	53	49	0.25	0.32	39	64	58	0.30	0.33	57	67
KUHN	34	0.10	0.50	19	57	52	0.25	0.37	42	70	58	0.30	0.32	56	67
MOLEC	44	0.08	0.32	45	46	56	0.25	0.30	59	55	50	0.24	0.34	43	58
EIMCL	29	0.06	0.40	17	47	48	0.24	0.31	42	59	58	0.30	0.33	58	65
ROSE2	28	0.05	0.48	15	49	49	0.25	0.33	38	66	59	0.30	0.33	62	64
HELL3	34	0.05	0.39	21	56	52	0.24	0.33	48	61	56	0.26	0.31	54	62
FAUPL	27	0.04	0.40	16	45	49	0.25	0.36	42	63	59	0.31	0.33	61	64
KYTDO	29	0.04	0.46	14	53	48	0.22	0.34	38	67	58	0.29	0.32	58	65
CHOBET	33	0.02	0.44	14	64	50	0.21	0.35	41	67	57	0.26	0.32	57	64
CHOU5	27	0.01	0.43	11	55	48	0.25	0.34	36	69	59	0.30	0.34	62	64
FASBET	26	0.00	0.47	9	54	48	0.25	0.36	36	69	58	0.29	0.32	65	60
MEEK	26	0.00	0.40	16	43	48	0.22	0.31	46	56	57	0.30	0.34	64	58

Continued on next page

Table 1 (Continued)

Scale	Membrane					α-class					β-class				
	Q_3	C_α	C_c	$Q_{3\alpha}$	Q_{3c}	Q_3	C_α	C_c	$Q_{3\alpha}$	Q_{3c}	Q_3	C_β	C_c	$Q_{3\beta}$	Q_{3c}
NNEIG	25	−0.02	0.51	9	50	49	0.23	0.34	39	67	59	0.31	0.33	65	61
HELL2	36	−0.03	0.33	31	44	59	0.27	0.30	65	50	49	0.22	0.33	45	56
CHOU6	27	−0.05	0.47	10	56	49	0.24	0.35	38	68	57	0.28	0.31	54	68
ARGOS	29	−0.05	0.41	15	51	50	0.18	0.31	47	58	56	0.29	0.32	60	58
CHOTH2	25	−0.08	0.47	8	53	49	0.22	0.34	39	66	58	0.29	0.33	57	65
TYLOR	22	−0.09	0.43	6	47	48	0.23	0.32	41	61	58	0.30	0.34	64	60
JANIN	25	−0.12	0.51	7	53	50	0.21	0.34	42	66	57	0.28	0.32	57	64
VHEBL	22	−0.13	0.44	5	49	47	0.21	0.32	38	63	56	0.24	0.31	59	60
HOPPW	19	−0.16	0.38	5	41	48	0.22	0.32	44	58	57	0.28	0.33	66	56
KUNTZ	19	−0.16	0.40	5	41	52	0.25	0.31	54	54	55	0.26	0.31	69	48
LUND	18	−0.19	0.37	4	41	49	0.21	0.32	47	55	56	0.28	0.34	65	54
LEVIT	19	−0.20	0.35	4	41	48	0.20	0.31	45	54	55	0.27	0.32	64	55
ENGEL	19	−0.21	0.40	5	41	50	0.22	0.32	50	53	54	0.26	0.32	61	53

decreasing correlation coefficient C_α for predicting α-helical segments in membrane proteins. Performance measures for the β-sheet conformation in membrane proteins and α-class proteins and for the α-helix in β-class proteins are not reported, because only a small percentage of residues from each testing set of proteins are in such a conformation.

Prediction results for the segments in the α-helix conformation in five integral membrane proteins are better with our scale (JTL) than ones obtained after application of other secondary structure prediction programs. For instance, the Garnier and Robson program [3] applied at the same set of membrane proteins results in $C_\alpha = 0.23$ and $Q_3 = 58\%$. It is not surprising that the technique directed at predicting soluble protein structures gives poor results when applied to membrane proteins. It was even argued that all techniques trained on a data base of soluble protein structures are inappropriate as predictors of membrane proteins folding motifs [25]. While this belief may be supported for the case of some algorithms used to predict secondary structure, our results show that more careful analysis of the data base of soluble protein structures can increase accuracy of predicting membrane protein structures as well.

Nine of the top ten scales in table 1 have in common that protein specific properties are taken into account in all of them. These properties are originally derived from a statistical analysis of the data base of known structures of globular soluble proteins. Our own analysis of soluble protein structures used these properties to calculate preference functions that are then incorporated into a simple secondary structure prediction algorithm (see section 2). The success in predicting α-helix conformation both in membrane proteins and in α-class proteins, when some of these scales are used (table 1), indicates similarity of folding mechanisms leading to α-helix formation in soluble proteins and in integral membrane proteins. The only exception among the top ten scales in the table 1 is the SCHERA scale of helix pro-

pagation parameters that are, however, closely related to helix propensities derived from globular protein frequencies [26].

The hydrophobicity scales, often used to predict membrane buried segments of membrane proteins [9,16], are respectively 37th (FAUPL), 38th (KYTDO), 54th (LEVIT) and 55th (ENGEL) in the performance measure (C_α) for predicting helical segments in membrane proteins. Apparently, hydrophobicity alone is not enough to predict α-helix segments with high accuracy, even if methods used to measure hydrophobicity of amino acids vary greatly from one author to another. Indeed, each of the ten best scales uses other properties of amino acid residues in a protein environment in addition to hydrophobicity.

The dependence of Chou–Fasman preferences (JTL and FASTUR scale) on inter- and intra-molecular forces and on steric effects is well known [27]. The MIJER scale includes not only hydrophobic energy but also average contributions of electrostatic, hydrogen bonds and van der Waals energies to interresidue contact energies. The ROSE1 scale is based on the area a residue buries upon folding, i.e. includes steric, entropic and hydrophobic contributions to the conformational free energy. Both ROSE1 and WERSC scale are based on the idea that partitioning of a given residue between the inside and outside of the protein, assessed as the fraction of the residue population that is buried, ought to be a better measure of residue "hydrophobicity" than solution measurements.

Neither α-helix propensities (with exception of the ROBSON scale from 1971) nor β-sheet propensities are among the best conformational parameters for predicting helical segments in membrane proteins. The best scale (JTL) is constructed by us as an average of the Chou–Fasman preferences [28] for α-helix and β-sheet structures for each coded amino acid. It is of interest that the scale of turn preferences (FASTUR) and the chain flexibility scale (KARPLU) proposed for predicting turns are among the best predictors of helical segments in membrane proteins. To summarize, the data from table 1 for membrane proteins indicate that helix propagation in a membrane environment may require rather high chain flexibility and high potential for accessible surface area loss, but the same or similar requirements are likely to exist for nucleation of helices in α-class proteins too.

For twelve α-class proteins, six of the ten abovementioned scales (JTL, SCHERA, FASTUR, ROBSON, RICH1, CHOU3) are the best predictors of helix conformation. Not surprisingly, three scales for α-helical propensities: RICH1, FASALF and CHOU1 are also among ten best predictors of helix conformation in the α-class proteins.

For sixteen β-class proteins the best predictors of β-sheet conformation are solution hydrophobicity scales based upon a comparison of solubilities in aqueous and nonaqueous solvents (FAUPL, NNEIG, KRIGK, JONES, PONNU, EIMCL and TYLOR). Some scales calculated empirically from X-ray elucidated coordinates, such as GIBRAT, GUY and ROSE2, are also good predictors of secondary structure segments in β-class proteins, but in general the accuracy of prediction for such proteins is lower than for α-class and membrane proteins. For instance

maximal Q_3 of 59% for the optimal matching hydrophobicity scale (SWEET) of Sweet and Eisenberg [29] and for several of the abovementioned scales is eight points lower than maximal Q_3 for membrane and α-class proteins. Nevertheless, the evaluation of hydrophobicity scales on the testing set of β-class proteins indicated that β-sheet creation may be stimulated among residues and primary structure segments that favor partitioning in nonaqueous solvents if steric hindrance for joining such segments can be overcome.

Helix packing [30,31] and β-sheet packing [32] requirement for hydrophobic interactions is being studied and will be reported elsewhere. The most important conclusion from the analysis in the present paper may well be that the prediction of α-helix conformation in membrane proteins is extremely sensitive to the choice of twenty conformational parameters. This fact can be used not only to improve structure prediction for such proteins, but also to improve their identification when only primary sequence is known. However, it remains to be seen, whether such a conclusion holds only for energy transducing and signal transducing membrane proteins, as it does for those used in this study.

Acknowledgement

This work was supported by the Croatian Ministry of Science Grant 1-03-171.

Appendix. Index to the scales used in this article

JTL: It is the scale that we formed by averaging for each amino acid type the Chou and Fasman preferences [28] for α-helix and β-sheet structure. These preferences are derived from the observed frequencies in 29 soluble proteins.

SCHERA: Wojcik et al. [33]. According to the Zimm–Bragg formalism [34] experimental values for the thermodynamic parameter s are related to the ease of helix propagation.

FASTUR: Chou and Fasman turn preferences [35] derived from the observed frequencies in soluble proteins.

ROBSON: The scale of α-helical propensities derived from the observed frequencies [36] in soluble proteins.

MIJER: Miyazawa and Jernigan [37]. An average contact energy multiplied with average number of residue–residue and residue–solvent contacts. This scale is based on crystallographic data for 42 soluble proteins.

ROSE1: Rose et al. [38]. Mean solvent accessible surface area loss A for each amino acid type is computed for 23 soluble proteins.

KARPLU: Karplus and Schultz [39]. Experimentally determined chain flexibility scale for residues with no rigid neighbors. The data base consisted of 31 soluble proteins.

RICH1: α-helix preferences for middle helix regions: five-point averages of the values in the Richardson and Richardson paper [40]. The data base consisted of 45 globular soluble proteins.

CHOU3: Chou [41]. The helical preferences for 14 soluble proteins of $\alpha + \beta$ class.

WERSC: Wertz and Scheraga [42]. In the data base of 20 soluble proteins all residues are classified as being inside or outside. The scale is the ratio of interior X-residues to total X residues for each amino acid type X.

RICH2: α-helix preferences for middle helix regions: Middle column in the table 1 of Richardson and Richardson paper [40]. The data base consisted of 45 globular soluble proteins.

BULDG: Bull and Breese [43]. Experimental values for the slope of the surface tension relative to the concentration of the amino acid. These values are related to the free energy of transfer of the amino acid from solution to the surface.

DEBER: Deber et al. [44]. Mean membrane/aqueous domain occurrence ratio for residues within 10 transport proteins.

ABODR: Aboderin [45]. Mobilities of the amino acids on the Whatman No. 3 paper, using the apolar solvent system.

GUYM: Guy [46]. It is the average of four hydrophobicity scales, his own, Ponnuswamy et al., Meirovitch, and Wertz and Scheraga delta *G*.

CHART: Charton and Charton [27]. The polarizability parameters. The proline is added to this (incomplete) scale with a value of 0.00.

FASALF: Chou and Fasman preferences [28] for the α-helix structure. These conformational parameters are derived from the observed frequencies in twenty-nine soluble proteins.

HELL1: Hellberg et al. [47]. Descriptor scale z1 derived from a principal component analysis of a property matrix for the 20 coded amino acids. Mainly related to hydrophobicity.

CHOU2: Chou [41]. The β-sheet preferences for 15 soluble proteins of β class.

NEIL: O'Neil and DeGrado [48]. Helix formation parameters determined from the concentration dependence of helix formation for the peptides with different amino acids substituted in the guest position.

COHEN: Cohen and Kuntz [49]. Nonpolar area for residues in actual isolated β-sheets.

GIBRAT: Gibrat [50]. Partition coefficients of the amino acid residues between the interior and the exterior of a protein.

KRIDG: Krigbaum and Komoriya [51]. It is the list of interaction parameters for ethanol to water transfer of amino acid side chains.

CHOU4: Chou [41]. The α-helix preferences for 16 soluble proteins of α/β class.

PRIFT: Cornette et al. [2]. The scale that maximizes the amphipathic index for the 145 helices from 23 proteins.

SWEET: Sweet and Eisenberg [29]. Optimal matching hydrophobicity scale based on point mutations.

CHOU1: Chou [41]. The α-helix preferences for 19 soluble proteins of α-class.

JONES: Jones [52]. The hydrophobicity scale based on experimental data given by Nozaki and Tanford [53].

CHOTH1: Chothia [54]. Accessible surface area of individual residues, R, in the tripeptide Gly–R–Gly.

PONNU: Ponnuswamy et al. [55]. The surrounding hydrophobicity of residue X in a protein is computed as the sum of the Jones hydrophobicities of all of the other residues in the protein whose α-carbon atoms are within 8 Å of the α-carbon of residue X.

GUY: Guy [46]. A statistical hydrophobicity scale based on classifying residues of 19 proteins as lying in one of six layers, from the surface to the center of the protein.

KUHN: Kuhn and Leigh [8]. Membrane propensity scale based on the frequency of occurrence of the amino acid in transmembrane segments. The transmembrane assignments for residues from ten membrane proteins was used as the data base.

MOLEC: Molecular weight of amino acids.

EIMCL: Eisenberg and MacLachlan [56]. Computed solvation energy of each amino acid. The contribution from each atom is found as the product of its solvent accessible area [57] and its atomic solvation parameter. Solvation parameters are computed to give the best fit of the transfer energies to the scale of Fauchère and Pliška [58].

ROSE2: Rose et al. [38]. Mean fractional area loss is computed for each amino acid as it appears in 23 proteins. It is obtained from the ROSE1 values A by forming the ratio A/A_0, where A_0 is the solvent accessible surface area of the amino acid, X, in a standard state.

HELL3: Hellberg et al. [47]. Descriptor scale z3 derived from a principal component analysis of a property matrix for the 20 coded amino acids. Contains information from pK and ^{1}H, ^{13}C NMR variables.

FAUPL: Fauchère and Pliška [58]. This hydrophobic scale is based on octanol/water distribution measurements for all 20 N-acetyl-amino acid amides.

KYTDO: Kyte and Doolittle [7]. It is hydrophobicity scale that combines Wolfenden scale [59], the Chothia scale [54] and estimates based on the constituent parts of the side chain.

CHOBET: Chou [60]. The frequency of amino acids in the known β-sheets.

CHOU5: Chou [41]. The β-sheet preferences for 14 soluble proteins of $\alpha + \beta$ class.

FASBET: Chou and Fasman preferences [28] for the β-sheet structure. These conformational parameters are derived from the observed frequencies in 29 soluble proteins.

MEEK: Meek [61]. The scale based on the retention time measurements of 25 peptides in high-pressure liquid chromatography.

NNEIG: Cornette [2]. A dominant eigenvector of the nearest neighbor matrix based on Ponnuswamy [55] hydrophobicity scale.

HELL2: Hellberg et al. [47]. Descriptor scale z2 derived from a principal component analysis of a property matrix for the 20 coded amino acids. Contains information about size.

CHOU6: Chou [41]. The β-sheet preferences for 16 soluble proteins of α/β class.

ARGOS: Argos and Palau [62]. Conformational propensity parameters for middle regions of β-structures. The scale was determined from the Levitt and Greer [63] sample of 60 known protein structures.

CHOTH2: Chothia [54]. Proportion of residues 95% buried in 12 soluble proteins.

TYLOR: Taylor and Thorton [5]. Hydrophobicity was scored on a simple scale ranging from 2 to −2.

JANIN: Janin [64]. The scale is the ratio of buried to accessible molar fractions of each amino acid, as measured in 22 proteins.

VHEBL: von Heijne and Blomberg [65]. The free energy of transfer of a single residue in a polypeptide from a random coil conformation in an aqueous phase to the helix conformation in the nonpolar environment of membrane interior.

HOPPW: Hopp and Wood [66]. Hydrophilicity values assigned to the 20 amino acids commonly found in proteins. The Levitt scale [63] is adjusted so that the hydrophobicity profile would more successfully identify antigenic determinants in 12 proteins.

KUNTZ: Kuntz [67]. Hydration scale which measures the amount of water that does not freeze when an aqueous macromolecular solution is rapidly frozen.

LUND: Lundeen and Chance [68]. The estimated hydrophobic free energy gain when a side chain in a protein in random conformation is taken out of contact with aqueous solution.

LEVIT: Levitt [69]. Hydrophobicity scale calculated from solubilities of the amino acids in water and ethanol by Nozaki and Tanford [53] and supplemented with accessible surface area values given by Chothia [54].

ENGEL: Engelman et al. [9]. A polarity scale for identifying transmembrane helices that combines surface area and polar contributions arising from hydrogen bonding interaction.

References

[1] W. Kauzman, Adv. Protein Chem. 14 (1959) 1.

[2] J.L. Cornette, K.B. Cease, H. Margalit, J.L. Spouge, J.A. Berzofsky and C. DeLisi, J. Mol. Biol. 196 (1987) 659.

[3] J. Garnier, D.J. Osguthorpe and B. Robson, J. Mol. Biol. 120 (1978) 97.
[4] B. Busetta and M. Hospital, Biochim. Biophys. Acta 701 (1982) 111.
[5] W.R. Taylor and J.M. Thorton, J. Mol. Biol. 173 (1984) 487.
[6] D.G. Kneller, F.E. Cohen and R. Langridge, J. Mol. Biol. 214 (1990) 171.
[7] J. Kyte and R.F. Doolittle, J. Mol. Biol. 157 (1982) 105.
[8] L.A. Kuhn and J.S. Leigh Jr., Biochim. Biophys. Acta 828 (1985) 351.
[9] D.M. Engelman, T.A. Steitz and A. Goldman, Ann. Rev. Biophys. Biophys. Chem. 15 (1986) 321.
[10] D. Eisenberg, E. Schwarz, M. Komaromy and R. Wall, J. Mol. Biol. 179 (1984) 125.
[11] A. Ben-Naim, *Hydrophobic Interactions* (Plenum Press, New York, 1980).
[12] R.M. Sweet and D. Eisenberg, J. Mol. Biol. 171 (1983) 479.
[13] D. Juretić, B. Lee, N. Trinajstić and R.W. Williams, Biopolymers 33 (1993) 255.
[14] D. Juretić and R.W. Williams, J. Math. Chem. 8 (1991) 229.
[15] G.D. Rose, Nature 272 (1978) 586.
[16] J. Kyte and R.F. Doolittle, J. Mol. Biol. 157 (1982) 105.
[17] R.A. Fisher, *Statistical Methods for Research Workers* (Hafner, New York, 1973).
[18] B.W. Mathews, Biochim. Biophys. Acta 405 (1975) 442.
[19] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F.J. Meyer, J.R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, J. Mol. Biol. 112 (1977) 535.
[20] L.J. Stern, P.L. Ahl, T. Mart, T. Mogi, M. Dunach, S. Berkowitz, K.J. Rothschild and H.G. Khorana, Biochemistry 28 (1989) 10035.
[21] J. Deisenhofer, O. Epp, K. Mikki, R. Huber and H. Michel, Nature 318 (1985) 618.
[22] H. Michel, K.A. Weyer, I. Gruenberg, I. Dunger, D Oesterhelt and F. Lottspeich, EMBO J. 5 (1986) 1149.
[23] P.D. Roepe and H.R. Kaback, Biochemistry 28 (1989) 6127.
[24] R.P. Birge, Biochim. Biophys. Acta 1016 (1990) 293.
[25] B.A. Wallace, M. Cascio and L. Mielke, Proc. Natl. Acad. Sci. US 83 (1986) 9423.
[26] P.Y. Chou and G.D. Fasman, Biochemistry 13 (1974) 211.
[27] M. Charton and B.I. Charton, J. Theor. Biol. 102 (1983) 121.
[28] P.Y. Chou and G.D. Fasman, Annu. Rev. Biochem. 47 (1978) 251.
[29] R.M. Sweet and D. Eisenberg, J. Mol. Biol. 171 (1983) 479.
[30] K.-C. Chou, G. Nemethy and H.A. Scheraga, J. Phys. Chem. 87 (1983) 2869.
[31] K.-C. Chou, G. Nemethy and H.A. Scheraga, J. Am. Chem. Soc. 106 (1984) 3161.
[32] K.-C. Chou, L. Carlacci and G.G. Maggiora, J. Mol. Biol. 213 (1990) 315.
[33] J. Wojcik, K.-H. Altman and H.A. Scheraga, Biopolymers 30 (1990) 12.
[34] B.H. Zimm and W.K. Bragg, J. Chem. Phys. 31 (1959) 526.
[35] P.Y. Chou and G.D. Fasman, Adv. Enzymol. 47 (1978) 45.
[36] B. Robson and R.H. Pain, J. Mol. Biol. 58 (1971) 237.
[37] S. Miyazawa and R.L. Jernigan, Macromolecules 18 (1985) 534.
[38] G.D. Rose, A.R. Geselowitz, G.J. Lesser, R.H. Lee and M.H. Zehfus, Science 229 (1985) 834.
[39] P.A. Karplus and G.E. Schultz, Naturwissenschaften 72 (1985) 212.
[40] J.S. Richardson and D.C. Richardson, Science 240 (1988) 1648.
[41] P.Y. Chou, in: *Prediction of Protein Structure and Principles of Protein Conformation*, ed. G.D. Fasman (Plenum Press, New York, 1989) pp 549–586.
[42] D.H. Wertz and H.A. Scheraga, Macromolecules 11 (1978) 9.
[43] Bull, H.B. and K. Breese, Arch. Biochem. Biophys. 161 (1973) 665.
[44] C.M. Deber, C.J. Brandl, R.B. Deber, L.C. Hsu and X.K. Young, Arch. Biochem. Biophys. 251 (1986) 68.
[45] A.A. Aboderin, Int. J. Biochem. 2 (1971) 537.
[46] H.R. Guy, Biophys. J. 47 (1985) 61.
[47] S. Hellberg, M. Sjostrom and S. Wold, Acta Chem. Scand. B40 (1985) 135.

[48] K.T. O'Neil and W.F. DeGrado, Science 250 (1990) 646.
[49] F.E. Cohen and I.D. Kuntz, in: *Predictlon of Protein Structure and Principles of Protein Conformation*, ed. G.D. Fasman (Plenum Press, New York, 1989) pp. 647–705.
[50] J.F. Gibrat, Modelisation by computers of the 3-D Structure of Proteins, PhD Thesis, University of Paris VI, France (1986).
[51] W.R. Krigbaum and A. Komoriya, Biochim. Biophys. Acta 576 (1979) 204.
[52] D.D. Jones, J. Theor. Biol. 50 (1975) 167.
[53] Y. Nozaki and C. Tanford, J. Biol. Chem. 246 (1971) 2211.
[54] C. Chothia, J. Mol. Biol. 105 (1976) 1.
[55] P.K. Ponnuswamy, M. Prabhakaran and P. Manavalan, Biochim. Biophys. Acta 623 (1980) 301.
[56] D. Eisenberg and A.D. McLachlan, Nature (London) 319 (1986) 199.
[57] B. Lee and F.M. Richards, J. Mol. Biol. 55 (1971) 379.
[58] J. Fauchère and V. Pliška, Eur. J. Med. Chem. 18 (1983) 369.
[59] R. Wolfenden, L. Anderson, P.M. Cullis and C.C.B. Southgate, Biochemistry 20 (1981) 849.
[60] K.-C. Chou, Biophys. J. 48 (1985) 289.
[61] J.L. Meek, Proc. Natl. Acad. Sci. US 77 (1980) 1632.
[62] P. Argos and J. Palau, Int. J. Peptide Protein Res. 19 (1982) 380.
[63] M. Levitt and J. Greer, J. Mol. Biol. 114 (1977) 181.
[64] J. Janin, Nature 277 (1979) 491.
[65] G. von Heijne and C. Blomberg, Eur. J. Biochem. 97 (1979) 175.
[66] T.P. Hopp and K.R. Woods, Proc. Natl. Acad. Sci. US 78 (1981) 3824.
[67] I.D. Kuntz, J. Am. Chem. Soc. 93 (1971) 514.
[68] M. Lundeen and B. Chance, Biophys. J. 51 (1987) 693.
[69] M. Levitt, J. Mol. Biol. 104 (1976) 59.