SUPPLEMENTARY DATA

METHODS

Calculation of algorithms performance:

Sensitivity or true positive rate was calculated as: Sensitivity = TP/P = TP/(TP+FN).
Specificity or true negative rate was calculated as: Specificity = TN/N = TN/(FP+TN).
Precision or positive predictive value was calculated as: Precision = TP/(TP+FP)
Where P = Positives, N = Negatives, TP = True Positives, FP = False Positives, TN = True Negatives and FP = False Positives.

Table S1: Set of 29 different WT proteins for which 3D-structures are available and changes in solubility upon mutation and expression have been characterized, accounting a total of 129 variants. PDBs, mutations and the corresponding references are indicated. Experimental data are compared with PROSOII, SolPro and A3D predictions. Green cells correspond to correct predictions whereas red cells indicate wrong predictions. True positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) are indicated.

Figure S1**:** A3D prediction pipeline

Figure S2**:** The aggregation propensity of (a) Maltose Binding Protein (PDB: 4MBP) and (b) Ubiquitin (PDB: 1UBQ) are analysed using both sequence-based predictors and A3D Amino acids with positive scores are indicated in different colours for each predictor: AGGRESCAN (green), Zyggregator (blue), FoldAmyloid (red), Tango (orange) and A3D (purple). Black bars indicate buried residues exposing < 15 $Å^2$ (up) and < 30 $Å^2$ (down) to solvent. Areas with high-predicted aggregation propensity are labelled according to a red gradient and areas with high-predicted solubility according to a blue gradient. White areas are not predicted to influence aggregation.

Figure S3**:** Aggregation profiles of deoxyhemoglobin S and A. The aggregation propensity of the β-chains of deoxyhemoglobin S (solid line) and deoxyhemoglobin A (dashed line) were predicted with A3D, Aggrescan, Zyggregator, FoldAmyloid and Tango. The algorithm threshold (Tr) is shown as a solid red line. A3D is unique at pinpointing the large increase in aggregation propensity caused by the Glu6 to Val6 mutation. A3D predicts Val6 as the most aggregation-prone residue in deoxyhemoglobin S.

Figure S4: The oligomeric and monomeric forms of (a) TTR (PDB: 1TTA) and (b) SOD1 (PDB: 2C9V) were analysed with A3D. Color codes as in Figure S2. In both cases strong aggregation-prone regions in the monomer surface are protected at the interface of the native oligomer.

Figure S5: Comparison of A3D predictions for AcPDro2 (PDB: 1URR) in static (left) and dynamic (right) modes. The exposed aggregation-prone region shown in the dynamic analysis corresponds essentially to strand S5 and the preceding loop (Y89-I96) and strand S2 (W38).

Table S1

PROTEINS			Exp Data	1D-SEQ Predictions		AGGRESCAN 3D	References
Protein	PDB	Mutations		PROSOII score	SolPro score		
RNAse SA	1RGG:A	Thr76Ala	Increased	TP	TP	FN	[1]
		Thr76Arg	Increased	FN	TP	TP	
		Thr76Asn	Increased	TP	TP	TP	
		Thr76Asp	Increased	TP	TP	TP	
		Thr76Cys	Decreased	FP	FP	TN	
		Thr76Glu	Increased	FN	TP	TP	
		Thr76Gly	Increased	TP	FN	FN	
		Thr76His	Increased	TP	TP	TP	
		Thr76Ile	Decreased	FP	FP	TN	
		Thr76Leu	Decreased	FP	FP	TN	
		Thr76Lys	Increased	TP	TP	TP	
		Thr76Met	Decreased	TN	FP	TN	
		Thr76Phe	Decreased	FP	TN	TN	
		Thr76Pro	Decreased	TN	FP	TN	
		Thr76Ser	Increased	TP	TP	TP	
		Thr76Trp	Decreased	FP	FP	TN	
		Thr76Val	Decreased	FP	FP	TN	
A beta 42	1Z0Q:A	Phe19Ala	Increased	TP	TP	TP	[2]
		Phe19Arg	Increased	TP	FN	TP	
		Phe19Asn	Increased	TP	TP	TP	
		Phe19Asp	Increased	TP	TP	TP	
		Phe19Cys	Increased	TP	TP	TP	
		Phe19Gln	Increased	TP	TP	TP	
		Phe19Glu	Increased	TP	TP	TP	
		Phe19Gly	Increased	TP	TP	TP	
		Phe19His	Increased	TP	TP	TP	
		Phe19Ile	Decreased	FP	FP	TN	
		Phe19Leu	Increased	TP	TP	TP	
		Phe19Lys	Increased	TP	TP	TP	
		Phe19Met	Increased	TP	TP	TP	
		Phe19Pro	Increased	TP	TP	TP	
		Phe19Ser	Increased	TP	TP	TP	
		Phe19Thr	Increased	TP	TP	TP	
		Phe19Trp	Increased	TP	TP	TP	
		Phe19Tyr	Increased	TP	TP	TP	
		Phe19Val	Increased	TP	TP	TP	
Hemoglobin	2D60:B	Glu6Leu	Decreased	TN	TN	TN	[3]
		Glu6Phe	Decreased	TN	TN	TN	
		Glu6Trp	Decreased	TN	TN	TN	
A beta 42	1Z0Q:A	Ile32Ser	Decreased	FP	FP	TN	[4]
		Leu34Pro	Increased	TP	TP	TP	
		Phe19Ser // Leu34Pro	Increased	TP	TP	TP	
		Val12Ala // Ile32Thr // Leu34Pro	Increased	TP	TP	TP	
		Val12Glu // Val18Glu // Met35Thr // Ile41Asn	Increased	TP	TP	TP	
		Phe4Ile // Ser8Pro // Val24Ala // Leu34Pro	Increased	TP	TP	TP	
		His6Gln // Val12Ala // Val24Ala // Ile32Met // Val36Gly	Increased	FN	TP	TP	
Human HIV type 1 integrase	1BIZ:A	Trp131Ala	Increased	FN	FN	TP	[5]
		Val165Lys	Increased	FN	TP	TP	
FOP Protein	2D68 A	Leu104Met	Decreased	FP	FP	TN	[6]
		Leu127Met	Decreased	TN	TN	TN	
		Lys63Met	Decreased	TN	FP	TN	
		Lys69Met	Decreased	TN	FP	TN	
		Val74Phe	Decreased	TN	FP	TN	
		Leu87Met	Decreased	TN	FP	TN	
		Thr90Ala	Decreased	TN	FP	TN	
		Glu97Ala	Decreased	TN	TN	TN	
		Thr90Ala // Glu97Ala	Decreased	TN	FP	TN	
Human alpha-1 proteinase inhibitor	8API:A	Met358Leu	Decreased	TN	FP	TN	
		Thr345Leu // Met358Arg	Decreased	TN	FP	FP	
		Met351Glu // Met358Arg	Increased	FN	TP	TP	

Human Interleukin 1 Beta	9ILB:A	Leu10Asn	Decreased	TN	TN	TN	
		Leu10Asp	Decreased	TN	TN	TN	
		Lys97Arg	Increased	FN	FN	TP	
		Lys97Gly	Decreased	FP	TN	TN	
		Lys97Val	Decreased	FP	TN	TN	
Plasmodium falciparum dihydrofolate reductase	1J3I:A	Lys27Glu	Increased	TP	FN	TP	[7]
		Tyr35Gln // Phe37Arg	Increased	FN	TP	TP	
		Tyr35Gly // Phe37Leu	Increased	TP	TP	TP	
		Tyr35Leu // Phe37Arg	Increased	TP	TP	FN	
HIV integrase	1BIZ:A	Lys185Asn	Decreased	TN	TN	TN	[8]
		Lys185Asp	Decreased	TN	TN	TN	
		Lys185Ile	Decreased	TN	TN	TN	
		Lys185Leu	Decreased	TN	TN	TN	
		Lys185Phe	Decreased	TN	TN	TN	
		Lys185Val	Decreased	TN	TN	TN	
GP24	1YUE:A	Glu89Ala // Glu90Ala	Decreased	TN	FP	TN	[9]
CD58	1CI5:A	Ser1Phe // Lys9Val // Gln21Val // Lys58Val // Ser85Thr // Gly93Leu	Decreased	TN	TN	TN	[10]
Maltose Binding Protein	1JW4:A	Trp232Glu	Decreased	TN	FP	TN	[11]
		Tyr242Glu	Decreased	FP	FP	TN	
		Ile317Glu	Decreased	FP	FP	TN	
		Gly32Asp Ile33Pro	Decreased	FP	FP	FP	
Type S1 dihydrofolate reductase	2W9T:A	Asn48Glu	Increased	TP	FN	TP	[12]
		Asn130Asp	Increased	FN	TP	TP	
		Asn48Glu // Asn130Asp	Increased	FN	TP	TP	
Potassium channel KcsA	3EFF:K	Trp26Glu // Val93Glu // Leu24Asp // Leu81Arg // Leu116Arg	Increased	FN	FN	TP	[13]
Human translation initiation factor eIF2a	1Q8K:A	Ile27Gln // Ile46His // Val71Lys	Increased	TP	FN	TP	[14]
HIV type 1 integrase	1B9D:A	Phe185Lys	Increased	TP	FN	TP	[15]
Human apolipoprotein E C-terminal domain	2L7B:A	Phe257Ala // Trp264Arg // Val269Ala // Leu279Gln // Val287Glu	Increased	TP	TP	TP	[16]
Moloney murine leukemia virus reverse transcriptase	4MH8:A	Leu435Lys	Increased	TP	FN	TP	[17]
SIV integrase	1C6V:A	Phe185His	Increased	TP	TP	TP	[18]
Human apolipoprotein D	2HZR:A	Trp99His // Ile118Ser // Leu120Ser	Increased	TP	TP	TP	[19]
Cholera toxin A1 subunit	1S5F:A	Phe132Ser	Increased	TP	FN	TP	[20]
HhaI methyltransferase	2C7O:A	Val213Ser	Increased	TP	TP	TP	[21]
CD58	1CI5:A	Phe1Ser // Val9Lys // Val21Gln // Val58Lys // Thr85Ser // Leu93Gly	Increased	TP	TP	TP	[22]
Catalytic domain of beta4gal-T1	4L41:C	Ala155Glu // Asn160Lys // Met163Thr // Ala168Thr // Thr242Asn // Asn255Asp // Ala259Thr	Increased	TP	FN	TP	[23]
Human leptin	1AX8:A	Trp100Glu	Increased	TP	FN	TP	[24]
		Trp100Gln, Trp138Gln	Increased	TP	FN	TP	[25]
		His97Ser // Trp100Gln // Ala101Thr // Gly112Glu // Met136Ile // Trp138Gln // Gly145Glu	Increased	TP	TP	TP	
Hemoglobin	1DXT:D	Ser9Cys	Decreased	TN	FP	TN	[26]
		Glu6Lys	Decreased	FP	FP	TN	[27]

VL	2BX5:A	Arg24Asp	Increased	TP	TP	TP	[28]
		Ser26Asp	Increased	TP	TP	TP	
		Gln27Asp	Increased	FN	TP	FN	
		Ser28Asp	Increased	TP	TP	TP	
		Ile29Asp	Increased	FN	TP	TP	
		Ser30Asp	Increased	FN	TP	TP	
		Ser31Asp	Increased	FN	TP	TP	
		Tyr32Asp	Increased	FN	TP	TP	
		Asn34Asp	Increased	FN	TP	TP	
		Tyr49Asp	Increased	TP	TP	TP	
		Ala50Asp	Increased	TP	FN	TP	
		Ala51Asp	Increased	TP	TP	TP	
		Ser52Asp	Increased	FN	TP	TP	
		Ser53Asp	Increased	FN	TP	TP	
		Gln55Asp	Increased	TP	TP	TP	
		Ser56Asp	Increased	TP	FN	TP	
VH	3UPC	Gly26Asp	Increased	FN	TP	TP	
		Thr28Asp	Increased	FN	FN	TP	
		Ser30Asp	Increased	TP	FN	TP	
		Ser31Asp	Increased	FN	FN	TP	
		Tyr32Asp	Increased	FN	FN	TP	
		Als33Asp	Increased	TP	FN	TP	
		Gln39Asp	Increased	TP	FN	FN	
		Ala40Asp	Increased	TP	FN	FN	
		Ala50Asp	Increased	FN	TP	TP	
		Ser52Asp	Increased	FN	TP	TP	
		Gly53Asp	Increased	FN	FN	TP	
		Ser54asp	Increased	FN	FN	TP	
		Gly55Asp	Increased	FN	FN	TP	

References

[1] Trevino SR, Scholtz J, Pace C. 2007. J Mol Biol 366(2):449-460.

[2] de Groot N, Aviles F, Vendrell J, Ventura S. 2006. FEBS J 273(3):658-668.

[3] Adachi K, Konitzer P, Kim J, Welch N, Surrey S. 1993. J Biol Chem 268:21650-21656.

[4] Fisher A, Kim W, DeLisa M. 2006. Protein Sci 15(3):449-58.

[5] Idicula-Thomas S, Balaji PV. 2005. Protein Sci 14(3):582-592.

[6] Sim J, Sim T. 1999. Journal of Molecular Catalysis B: Enzymatic 6(3):133-143.

[7] Japrung D, Chusacultanachai S, Yuvaniyama J, Wilairat P, Yuthavong Y. 2005 Protein Eng Des Sel , 18(10):457-64.

[8] Maxwell KL, Mittermaier AK, Forman-Kay JD, Davidson AR. 1999. Protein Sci 8(9):1908-1911.

[9] Boeshans K, Liu F, Peng G, Idler W, Jang S, Marekov L, Black L, Ahvazi B. 2006. Protein Expr Purif 49(2):235-43.

[10] Sun ZYJ, Dotsch V, Kim M, Li J, Reinherz EL, Wagner G. 1999. EMBO J 18(11):2941-9.

[11]Wigley WC, Stidham RD, Smith NM, Hunt JF, Thomas PJ. 2001. Nature Biotech 19:131-136.

[12] Dale GE, Broger C, Langen H, D'Arcy A, Stuber D. 1994. Protein Eng 7:933–939.

[13] Slovic AM, Kono H, Lear JD, Saven JG, DeGrado WF. 2004. Proc Natl Acad Sci USA 101:1828–1833.

[14] Ito T, Wagner G. 2004. J Biomol NMR 28:357–367.

[15] Jenkins TM, Hickman AB, Dyda F, Ghirlando R, Davies DR, Craigie R. 1995. Proc Natl Acad Sci USA 92:6057–6061.

[16] Fan D, Li Q, Korando L, Jerome WG, Wang J. 2004. Biochemistry 43:5055–5064.

[17] Das D, Georgiadis MM. 2001. Protein Sci 10:1936–1941.

[18] Li Y, Yan Y, Zugay-Murphy J, Xu B, Cole JL, Witmer M, Felock P, Wolfe A, et al. 1999. Acta Crystallogr D Biol Crystallogr 55:1906–1910.

[19] Nasreen A, Vogt M, Kim HJ, Eichinger A, Skerra A. 2006. Protein Sci 15:190– 199.

[20] Agren L, Norin M, Lycke N, Lowenadler B. 1999. Protein Eng 12:173–178.

[21] Daujotyte D, Vilkaitis G, Manelyte L, Skalicky J, Szyperski T, Klimasauskas S. 2003. Protein Eng 16:295–301.

[22] Sun ZY, Dotsch V, Kim M, Li J, Reinherz EL, Wagner G. 1999. EMBO J 18:2941–2949.

[23] Malissard M, Berger EG. 2001. Eur J Biochem 268:4352– 4358.

[24] Zhang F, Basinski MB, Beals JM, Briggs SL, Churgay LM, Clawson DK, DiMarchi RD, Furman TC, et al.1997. Nature 387:206–209.

[25] Ricci MS, Pallito MM, Narhi LO, Boone T, Brems DN. 2006. Misbehaving Proteins: Protein (Mis)folding, aggregation, and stability. New York: Springer. pp 331–350.

[26] Bunn HF, Forget BG. 1986. Hemoglobin: Molecular, genetic and clinical aspects. Philadelphia: W. B. Saunders Co. pp 381–594.

[27] Charache S, Conley CL, Waugh DF, Ugoretz RJ, Spurrell JR. 1967. J Clin Invest 46:1795–1811.

[28] Dudgeon K, Rouet R, Kokmeijer I, Schofield P, Stolp J, Langley D, Stock D, Christ D. 2012. Proc Natl Acad Sci USA.109:10879-10884.

Figure S1

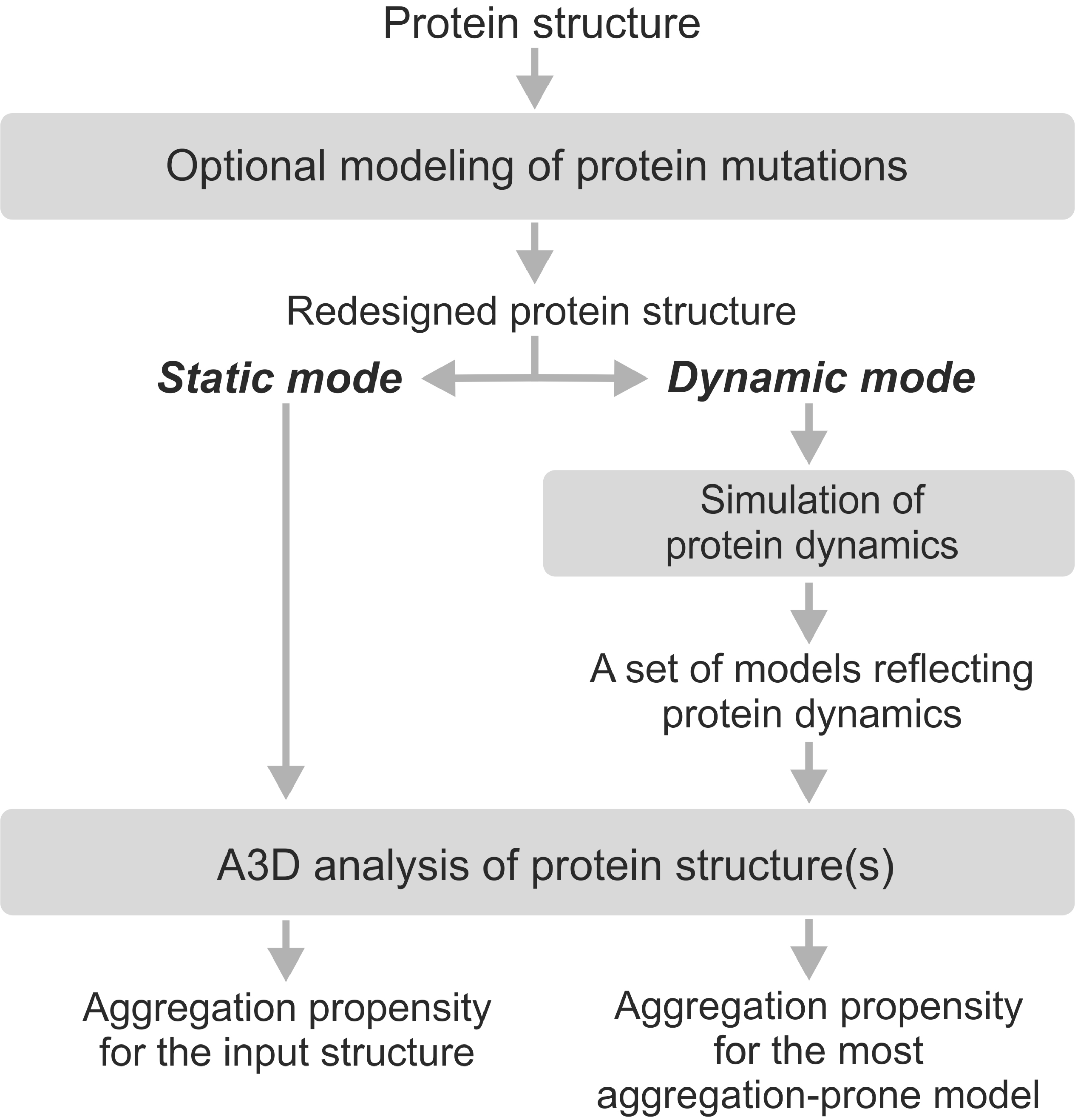
Protein structure
Optional modeling of protein mutations
Redesigned protein structure
Static mode
Dynamic mode
Simulation of protein dynamics
A set of models reflecting protein dynamics
A3D analysis of protein structure(s)
Aggregation propensity for the input structure
Aggregation propensity for the most aggregation-prone model

Figure S2

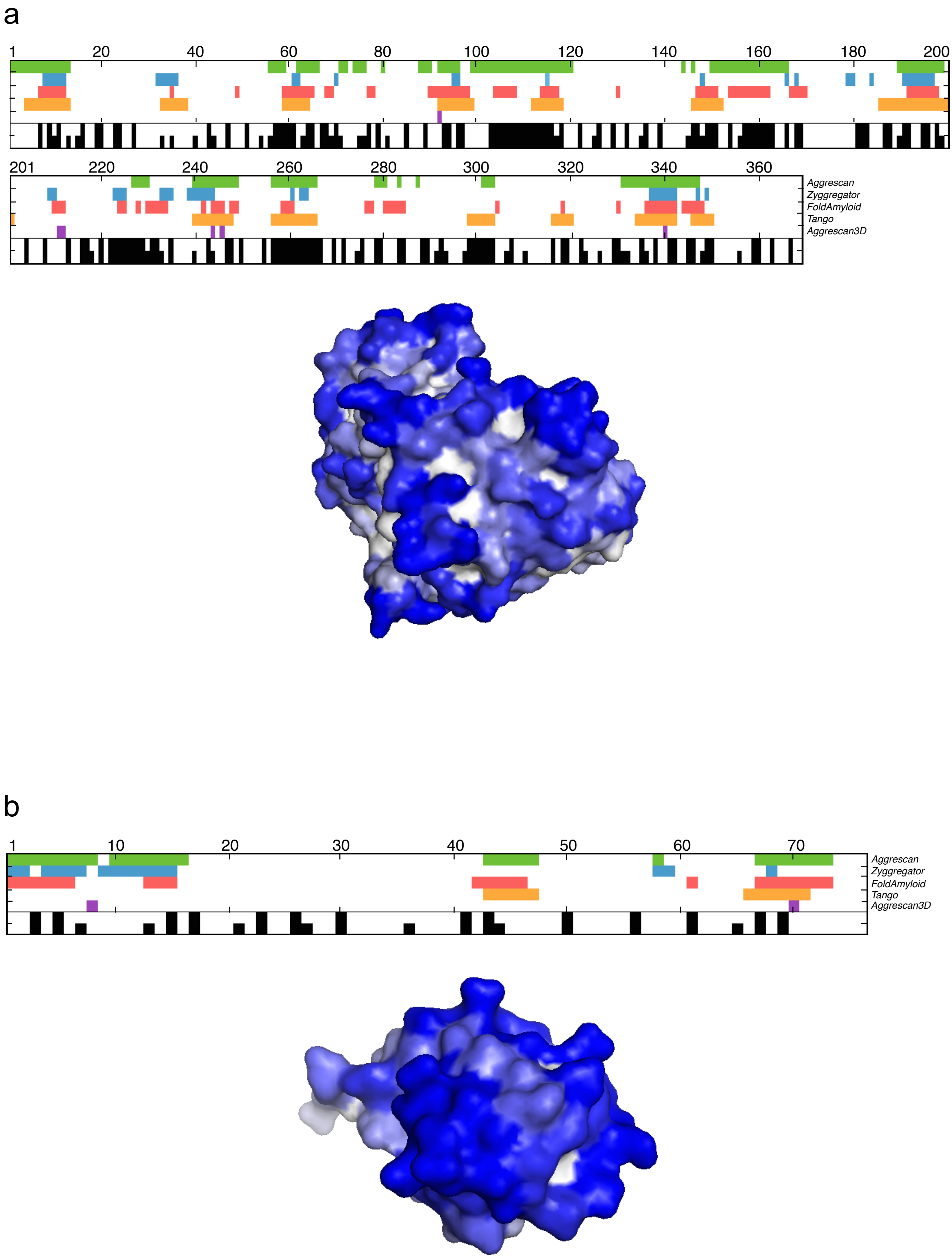
a
1
20
40
60
80
100
120
140
160
180
200
201
220
240
260
280
300
320
340
360
Aggrescan
Zyggregator
FoldAmyloid
Tango
Aggrescan3D
b
1
10
20
30
40
50
60
70
Aggrescan
Zyggregator
FoldAmyloid
Tango
Aggrescan3D

Figure S3

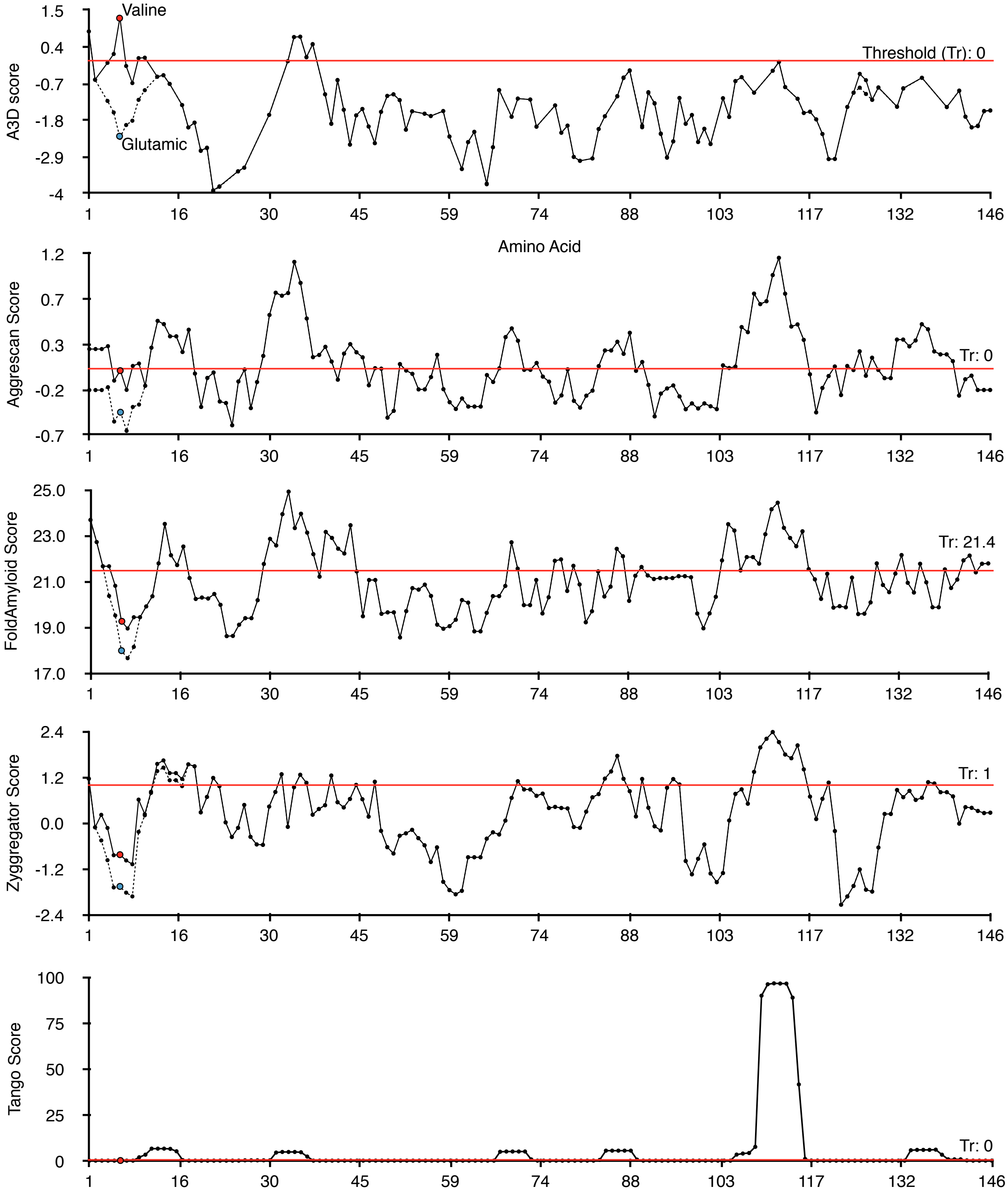
A3D score
1.5
0.4
-0.7
-1.8
-2.9
-4
Valine
Glutamic
Threshold (Tr): 0
1
16
30
45
59
74
88
103
117
132
146
Amino Acid
Aggrescan Score
1.2
0.7
0.3
-0.2
-0.7
Tr: 0
FoldAmyloid Score
25.0
23.0
21.0
19.0
17.0
Tr: 21.4
Zyggregator Score
2.4
1.2
0.0
-1.2
-2.4
Tr: 1
Tango Score
100
75
50
25
0
Tr: 0

Figure S4

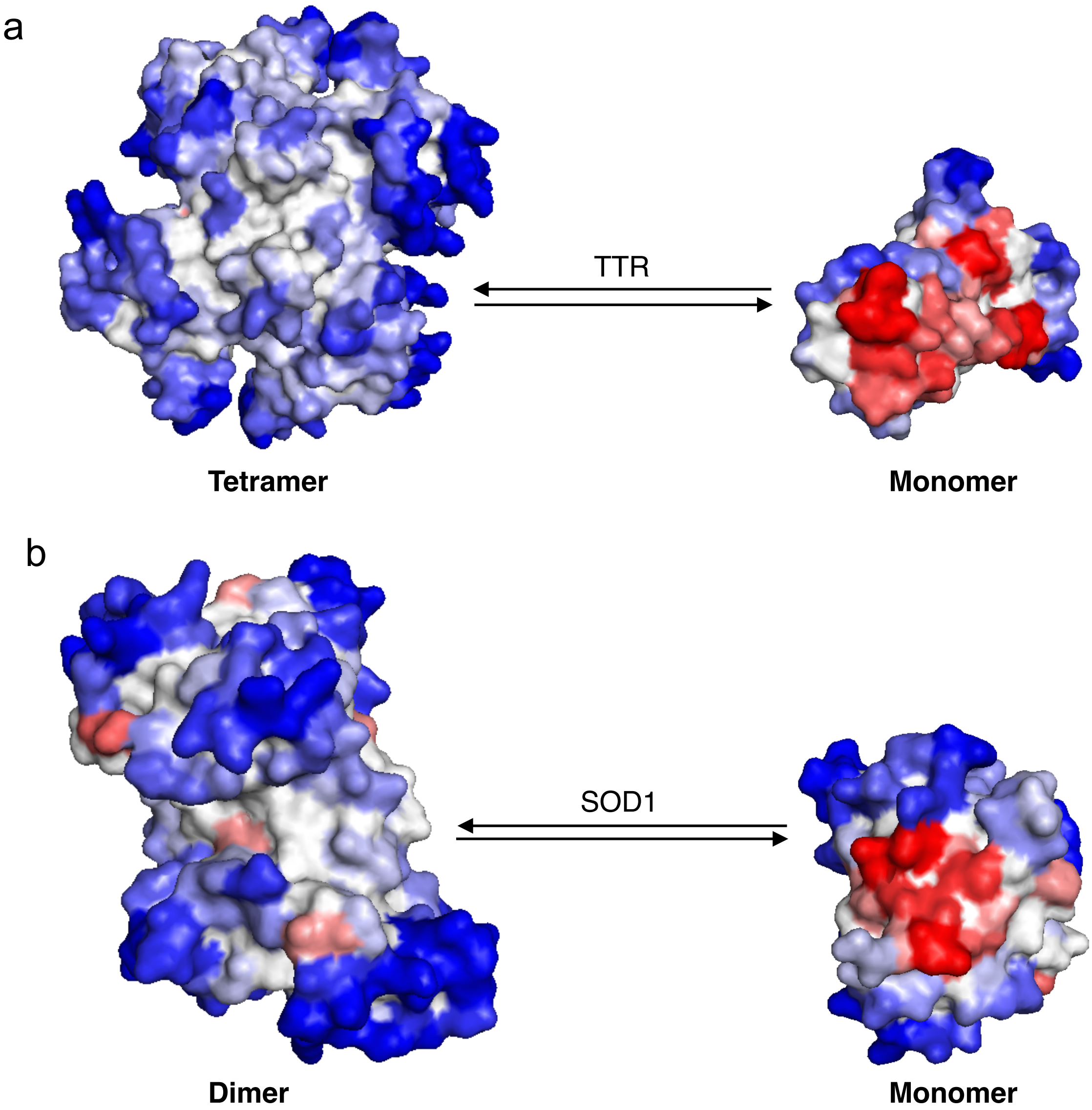

Figure S5