Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information

Are crystallographic B-factors suitable for calculating protein conformational entropy?

Octav Caldararu,[1] Rohit Kumar[3], Esko Oksanen,[2,3]

Derek T. Logan[3] & Ulf Ryde [1]*

[1] Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124, SE-221 00 Lund, Sweden

[2] European Spallation Source Consortium, P. O. Box 176, SE-221 00 Lund, Sweden

[3] Department of Biophysical Chemistry, Centre for Molecular Protein Science, Lund University, Chemical Centre, P. O. Box 124, SE-221 00 Lund, Sweden

Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se,

Tel: +46 – 46 2224502, Fax: +46 – 46 2228648

2019-07-17

Table S1. Data collection and refinement statistics for the room temperature galectin-3C structures. Statistics for the highest-resolution shell are shown in parentheses.

ligand	R	S
PDB code	6RHL	6RHM
station	BioMAX	BioMAX
wavelength [Å]	0.6525	0.6525
unit cell (Å)	a = 37.18 b = 58.40 c = 63.93	a = 37.29 b = 58.01 c = 63.81
space group	$P2_12_12_1$	$P2_12_12_1$
resolution range [Å]	29.20–1.29 (1.34–1.29)	31.91–1.59 (1.65–1.59)
completeness [%]	99.7 (98.1)	99.7 (98.6)
total reflections	451 933 (43 655)	166 040(14367)
unique reflections	34 990 (3423)	19 001(1853)
CC1/2	0.998 (0.266)	0.997 (0.438)
multiplicity	12.9 (12.8)	9.4 (7.7)
R_{merge} [%]	0.106 (3.11)	0.164 (0.98)
mean I/σ(I)	10.7 (0.6)	6.6 (0.5)
Wilson B-factor [Å^2]	18.7	20.5
R_{model} (F) [%]	0.136 (0.300)	0.178 (0.382)
R_{free} (F) [%]	0.165 (0.300)	0.190 (0.395)
reflections used in refinement	34 935 (3388)	18 958 (1841)
and for R_{free}	1697 (173)	987 (102)
average B-factors [Å^2]	protein: 23.0 ligand: 40.2 solvent: 38.4	protein: 23.5 ligand: 33.2 solvent: 34.5
Ramachandran outliers [%]	0.0	0.0
rotamer outliers [%]	0.0	0.73
MolProbity clash score	1.68	1.30
bond length rmsd from ideal [Å]	0.011	0.006
bond angle rmsd from ideal [°]	1.15	0.98

Table S2. List of residues in alternative conformations in the various studied protein crystal structures.

galectin-3C		trypsin	lysozyme
100 K	300 K		
Pro117		Val31	Met1
Asn119		Asn48	Glu5
Arg129	Arg129	Gln50	Thr21
Ile134		Val53	Ser44
	Lys139	Leu67	Asn53
Pro140	Pro140	Ser84	Val57
	Asn141	Lys87	Asn68
	Ala142	Ser88	Asp72
Asn143	Asn143	Pro92	Arg80
Arg144		Met104	Met106
Leu147	Leu147	Ser113	
Gln150	Gln150	Arg117	
Arg168	Arg168	Ser120	
Arg169	Arg169	Ser122	
Ile171	Ile171	Thr125	
Lys176		Cys128	
Ser188		Ser130	
Glu183	Arg183	Cys136	
	Glu193	Thr149	
Ser194		Tyr151	
Lys196		Asp153	
Glu205		Lys156	
	Asp207	Ser166	
His208	His208	Ser167	
	Lys210	Ser170	
Asn222	Asn222	Pro173	
Lys226	Lys226	Gln175	
	Lys227	Gln192	
Ser232	Ser232	Ser195	
Lys233	Lys233	Cys201	
	Ile236	Ser217	
Ser237	Ser237	Cys232	
Ser244	Ser244	Ser236	
Met249	Met249	Gln240	
		Ile242	
		Ser244	

Table S3. Amber topology files for the two ligands of trypsin, benzamidine (BAM) and benzylamine (ABN), and for the two ligands bound to lysozyme, N-phenylglycinonitrile (PGN) and methylbenzylazide (MBA).

BAM.in

```
    0    0    2

Benzamidine, RESP charges
molecule.res
BAM   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.523   111.21       .0      .00000
   4  N1    nh    M    3   2   1     1.540   111.208  -180.000 -0.893800
   5  H1    hn    E    4   3   2     1.010    76.662    57.846  0.398300
   6  H8    hn    E    4   3   2     1.010   122.044   -59.392  0.398300
   7  C     ca    M    4   3   2     1.306    72.537  -174.335  0.638800
   8  N2    n2    S    7   4   3     1.306   119.310   -62.061 -0.835100
   9  H7    hn    E    8   7   4     1.010   109.484  -120.495  0.394400
  10  C1    ca    M    7   4   3     1.472   120.067   117.454 -0.198900
  11  C2    ca    M   10   7   4     1.382   120.497   -14.213 -0.105000
  12  H2    ha    E   11  10   7     1.080   119.982     2.450  0.143500
  13  C3    ca    M   11  10   7     1.380   120.865  -177.532 -0.134500
  14  H3    ha    E   13  11  10     1.080   119.986   179.660  0.136000
  15  C4    ca    M   13  11  10     1.356   120.126    -0.336 -0.117000
  16  H4    ha    E   15  13  11     1.080   120.018   178.737  0.134000
  17  C5    ca    M   15  13  11     1.373   119.601    -1.277 -0.134500
  18  H5    ha    E   17  15  13     1.080   120.019  -178.461  0.136000
  19  C6    ca    M   17  15  13     1.382   120.686     1.533 -0.105000
  20  H6    ha    E   19  17  15     1.080   120.013   179.863  0.143500


LOOP
   C6   C1

IMPROPER
   C1   N2    C   N1
   C2   C6   C1    C
   C3   C1   C2   H2
   C2   C4   C3   H3
   C3   C5   C4   H4
   C4   C6   C5   H5
   C5   C1   C6   H6

DONE
STOP
```

ABN.in

```
    0    0    2

Benzylamine, RESP charges
molecule.res
ABN   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.523   111.21       .0      .00000
   4  N     n3    M    3   2   1     1.540   111.208  -180.000 -0.916800
   5  H     hn    E    4   3   2     1.010    86.595    58.958  0.351800
   6  H1    hn    E    4   3   2     1.010    47.127   -61.773  0.351800
   7  C     c3    M    4   3   2     1.465   156.151   -74.931  0.203100
   8  HXT1  h1    E    7   4   3     1.090   109.430  -171.822  0.033700
   9  HXT2  h1    E    7   4   3     1.090   109.389    72.221  0.033700
  10  C1    ca    M    7   4   3     1.492   112.812   -49.781 -0.095300
  11  C2    ca    M   10   7   4     1.396   119.882   -71.502 -0.118000
  12  H2    ha    E   11  10   7     1.080   119.994    -0.501  0.138000
  13  C3    ca    M   11  10   7     1.382   120.324   179.498 -0.133500
  14  H3    ha    E   13  11  10     1.080   119.988  -179.547  0.131000
  15  C4    ca    M   13  11  10     1.382   120.203     0.440 -0.128000
  16  H4    ha    E   15  13  11     1.080   120.001  -178.408  0.130000
  17  C5    ca    M   15  13  11     1.368   119.768     1.594 -0.133500
  18  H5    ha    E   17  15  13     1.080   119.975   177.956  0.131000
  19  C6    ca    M   17  15  13     1.376   120.293    -2.019 -0.118000
  20  H6    ha    E   19  17  15     1.080   118.887  -179.582  0.138000


LOOP
   C6   C1

IMPROPER
    C   C2   C1   C6
   C3   C1   C2   H2
   C2   C4   C3   H3
   C3   C5   C4   H4
   C4   C6   C5   H5
   C5   C1   C6   H6

DONE
STOP
```

PGN.in

```
    0    0    2

N-phenylglycinonitrile, RESP charges
molecule.res
PGN   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.523   111.21       .0      .00000
   4  N3    n1    M    3   2   1     1.540   111.208  -180.000 -0.057000
   5  N2    n1    M    4   3   2     1.290    94.001  -110.139  0.221000
   6  N1    n2    M    5   4   3     1.306   178.485   104.932 -0.446200
   7  C13   c3    M    6   5   4     1.476   119.658    16.312  0.206100
   8  H131  h1    E    7   6   5     1.090   109.443    62.226  0.068700
   9  H132  h1    E    7   6   5     1.090   109.440  -177.740  0.068700
  10  C14   ca    M    7   6   5     1.527   109.470   -57.753 -0.122300
  11  C4    ca    M   10   7   6     1.392   119.156  -105.757 -0.113000
  12  H4    ha    E   11  10   7     1.080   120.003    -0.207  0.142000
  13  C7    ca    M   11  10   7     1.390   119.266   179.775 -0.076300
  14  C8    c3    3   13  11  10     1.520   119.479   179.696 -0.055800
  15  H81   hc    E   14  13  11     1.090   109.493  -179.984  0.046367
  16  H82   hc    E   14  13  11     1.090   109.483   -60.021  0.046367
  17  H83   hc    E   14  13  11     1.090   109.504    59.960  0.046367
  18  C11   ca    M   13  11  10     1.393   120.969    -0.413 -0.123000
  19  H11   ha    E   18  13  11     1.080   120.176  -179.334  0.134000
  20  C6    ca    M   18  13  11     1.393   119.823     0.679 -0.127000
  21  H6    ha    E   20  18  13     1.080   120.651   179.529  0.134000
  22  C5    ca    M   20  18  13     1.394   119.349    -0.441 -0.124000
  23  H5    ha    E   22  20  18     1.080   119.424   179.930  0.132000


LOOP
   C5  C14

IMPROPER
  C13   C4  C14   C5
   C7  C14   C4   H4
   C8   C4   C7  C11
   C7   C6  C11  H11
  C11   C5   C6   H6
   C6  C14   C5   H5

DONE
STOP
```

MBA.in

```
    0    0    2

Methylbenzylamide, RESP charges
molecule.res
MBA   INT  0
CORRECT     OMIT DU   BEG
  0.0000
   1  DUMM  DU    M    0  -1  -2     0.000      .0        .0      .00000
   2  DUMM  DU    M    1   0  -1     1.449      .0        .0      .00000
   3  DUMM  DU    M    2   1   0     1.523   111.21       .0      .00000
   4  CAF   ca    M    3   2   1     1.540   111.208  -180.000 -0.184500
   5  HAF   ha    E    4   3   2     1.080   120.792  -153.159  0.133500
   6  CAD   ca    M    4   3   2     1.398    48.252   -48.809 -0.097500
   7  HAD   ha    E    6   4   3     1.080   120.034  -105.288  0.129500
   8  CAC   ca    M    6   4   3     1.395   119.403    74.712 -0.169000
   9  HAC   ha    E    8   6   4     1.080   119.988   179.323  0.131000
  10  CAE   ca    M    8   6   4     1.392   119.927    -0.677 -0.097500
  11  HAE   ha    E   10   8   6     1.080   120.016  -179.355  0.129500
  12  CAG   ca    M   10   8   6     1.394   120.355     0.642 -0.184500
  13  HAG   ha    E   12  10   8     1.080   119.982   179.258  0.133500
  14  CAJ   ca    M   12  10   8     1.395   120.444    -0.755  0.135600
  15  NAI   nh    M   14  12  10     1.343   121.740   179.497 -0.718600
  16  HNAI  hn    E   15  14  12     1.010   119.986  -164.818  0.388800
  17  CAH   c3    M   15  14  12     1.476   123.670    15.193  0.178800
  18  HAH1  h1    E   17  15  14     1.090   109.413   -50.484  0.062200
  19  HAH2  h1    E   17  15  14     1.090   109.404  -167.188  0.062200
  20  CAB   c3    M   17  15  14     1.471   112.196    71.163  0.119800
  21  HAB1  h1    E   20  17  15     1.091    90.694    49.749  0.029700
  22  HAB2  h1    E   20  17  15     1.090    90.687     9.309  0.029700
  23  NAA   n3    M   20  17  15     1.139   178.529  -150.479 -0.912800
  24  HAA1  hn    E   23  20  17     1.010   109.484  -179.980  0.351300
  25  HAA2  hn    E   23  20  17     1.010   109.512   -59.993  0.351300


LOOP
  CAJ  CAF

IMPROPER
  CAJ  CAD  CAF  HAF
  CAF  CAC  CAD  HAD
  CAD  CAE  CAC  HAC
  CAG  CAC  CAE  HAE
  CAJ  CAE  CAG  HAG
  CAG  CAF  CAJ  NAI
  CAH  CAJ  NAI HNAI

DONE
STOP
```

Table S4: Calculated absolute conformational entropies of galectin-3C binding to the two diastereomeric ligands R and S. Results are given as $T\Delta S$, in kJ/mol. "uc" means unit cell.

Method		DH		QHA		BF-iso		BF-aniso		TLS		BF-100K+MD		BF-300K+MD	
		R	S	R	S	R	S	R	S	R	S	R	S	R	S
Crystallographic	100 K					22782	23251	22331	22779	3965	3936				
refinement	300 K					26363	26419	26378		24148	24215				
Standard MD		-3503±4	-3944±3	8732±18	8730±18	38949±140	38951±178	33707±53	33740±69			9607±79	11283±130	16630±43	16359±35
Crystal	1 uc	-5339±4	-5327±5	4062±7	4050±5	20733±62	20893±102	25097±23	25155±35			16713±15	16937±109	20083±425	19420±193
MD	2 uc	-5329±3	-5322±3	4008±5	4042±4	21017±52	21305±111	25166±17	25296±42			16552±82	16917±121	20121±220	19370±282
Ensemble	100 K	-4890±8	-4790±8	2835±2	3064±3	27309±227	27466±513	26544±155	26410±234			19381±362	19976±381		
refinement	300 K	-4305±14	-4301±13	3385±3	2901±2	34352±601	35088±553	33514±437	34105±441					22674±308	23561±517