# Appendix

## Algorithm for construction of crystallographic conformer libraries

An amino acid rotamer library is a compilation of commonly observed or low-energy side-chain conformations. Rotamer libraries are used in experimental structure determination, homology modeling, structure prediction, and protein design to suggest reasonable side-chain conformations. The use of rotamer libraries significantly improves the computational efficiency of these methods by approximating the continuous conformational freedom of amino acid side chains with a limited, discrete set of fixed conformations. Rotamer libraries are typically produced by culling structural data as side-chain $\chi$ angles from a database of X-ray crystal structures (1-3). Canonical combinations of $\chi$ angles for each amino acid type are determined statistically, and coordinates are built with standard geometry based on these angles to populate the libraries.

Blundell and coworkers suggested that amino acid side-chain libraries could be generated by clustering raw side-chain coordinates from the Protein Data Bank in Cartesian space (4). Each side chain in such a library would be a conformer with the same coordinates as some real side chain in the database, not a rotamer built with idealized geometry. This approach has two potential advantages over other methods. First, it allows conformations with non-ideal bond lengths and angles to be included in the library. Second, it allows the number of conformations to be expanded beyond the standard rotameric forms in a more natural way. For example, the Dunbrack libraries are typically expanded to a useful size by adding extra rotamers with $\chi$ angles plus or minus one standard deviation around the mean value; such expanded rotamers (and, in fact, the mean rotamers themselves) are not guaranteed to correspond to actual observed conformations. Conformer libraries generated by the method of Blundell and coworkers can be made as large as desired using an algorithm that judiciously adds conformers from the raw data set to the library.

**Raw PDB data**

The raw structural data we used to generate amino acid side-chain conformer libraries were taken from the high-quality protein crystal structure databases compiled by Richardson and Dunbrack. The union of these databases is a set of 1,011 unique structures with resolution 1.8 Å or better; the entire list is given below. For each structure, REDUCE (5) was used to add and optimize all hydrogen atoms, and to analyze and flip Asn, Gln and His side chains if necessary. The set of protein main chains taken from these structures was filtered so that no pair shared a sequence identity of more than 50% by pairwise BLAST (6) and every chain was at least 35 amino acids long. The side-chain conformations from these chains were filtered according to criteria suggested by Richardson and coworkers (3): 1: all atoms were present, and backbone atoms from the two adjacent residues were present; 2: no atom had a B-factor greater than 30, or an occupancy less than 1.0; 3: amide flips and hydrogen placement by REDUCE (5) were

unambiguous. In total, 149,813 residue side chains of all amino acid types met these criteria.

The molecular mechanics energy of each raw conformer was calculated using a DREIDING force field (7) with terms to describe bond stretches, angle bends, torsions, and nonbonded van der Waals interactions. The bond, angle, and torsion energies were each normalized by the number of relevant terms in the energy function; for example, the torsion energy of each conformer was divided by the number of torsions used to describe it in the force field. The total energy of each conformer was then normalized by subtracting the lowest observed energy of any conformer of the same amino acid type. The final normalized energy of each conformer is thus the excess energy above the minimum observed value. The two normalizations we employ allow the energies of different amino acid types to be compared more accurately. Inspection of histograms for each amino acid type indicated that a 10 kcal/mol normalized energy cutoff would only discard the most distant outliers for any amino acid type. Application of this cutoff eliminated 1,490 raw conformers, or about 1% of the total. A sample histogram for Leucine conformers is shown in Figure 5.

**Clustering of PDB data**

To our knowledge, Blundell and coworkers have not published the clustering algorithm they used to produce their libraries. The algorithm we used to generate our own libraries is based on their idea and is described here. In our conformer libraries, every side-chain conformation from the raw data set is assigned to exactly one group, called a cluster. Each cluster is represented by one of its members, the centroid, which is the member with coordinates closest to the average coordinates of all cluster members. For any particular assignment of raw conformers to clusters, the corresponding conformer library is simply a list of all of the cluster representatives and their coordinates.

**Clustering moves**

In our clustering algorithm, a locally optimal assignment of raw conformers to clusters is produced by manipulating the number of clusters and their membership via discrete clustering moves. We define three clustering moves: **Switch** allows a single raw conformer to leave one cluster and join another; **Merge** combines two clusters into one; **Split** allows a raw conformer to start a new cluster on its own. These moves are depicted in Figure 6. The decision whether to apply a particular move is made by comparing distances between pairs of raw conformers. The distance between a pair of conformers is measured by first performing a least-squares fit over their N, $C_\alpha$, $C_\beta$ , and C atoms, and then computing the RMSD between them using all side-chain heavy atoms and rotatable (hydroxyl and sulfahydryl) hydrogens. **Switch** is applied so that each raw conformer is a member of the cluster whose centroid is closest to it. **Merge** and **Split** are applied based on the value of the clustering parameter *p*: two clusters are merged if their centroids are within *p* of each other, whereas a conformer splits off and starts a new cluster if the closest centroid of any existing cluster is farther than *p* from it.

The clustering moves are applied in the following algorithm until the number of clusters converges.

1. Start with a small number of clusters (one was used in this work), and randomly assign a single raw conformer to each as the sole member and cluster representative.
2. Assign each raw conformer in the data set to the cluster whose centroid is closest.
3. While the number of clusters is not converged:
    a. Iteratively attempt to **Merge** pairs of clusters until no cluster can be further merged.
    b. For each conformer C:
        i. Measure the distance *d* between C and the centroid of every existing cluster.
        ii. If the distance *d* to the closest cluster centroid is greater than *p*, **Split** C off as its own cluster.
        iii. Else, **Switch** C to the closest cluster.
        iv. Recompute the centroid for every cluster that has changed membership.

The filtered side-chain data from the PDB were clustered to make both backbone-dependent and backbone-independent libraries. For the backbone-independent libraries, all of the data for each amino acid type were clustered at once, without regard to the backbone dihedrals ϕ and ψ of each raw conformer. Backbone-dependent libraries were generated by first breaking up the Ramachandran plot for the data into a grid of 10°×10° squares. The data in each grid square were clustered independently, essentially generating a separate rotamer library that can used for each ϕ/ψ bin.

The clustering parameter *p* controls the granularity of the conformer libraries; a smaller value of *p* results in more clusters and thus more conformers in the library. For example, our backbone-independent conformer library has 81 Arg conformers when $p$ = 1.6 Å and 852 when $p$ = 0.6 Å. When choosing a library, researchers must make a trade-off between more accurate modeling and faster calculations; the method we describe here allows us to generate many different libraries so that an appropriately sized library may be chosen for each potential application.

**1011 structures from the PDB used to generate our conformer libraries**

```
119l 153l 16pk 19hc 1a12 1a1i 1a1y 1a28 1a2p 1a2y 1a2z 1a3a 1a3c 1a4i 1a62 1a6m
1a73 1a7s 1a8d 1a8e 1a8i 1a8o 1a92 1aac 1aap 1aay 1aba 1ads 1agi 1agj 1ah7 1aho
1aie 1ajj 1ajs 1ak0 1ako 1akr 1aky 1amf 1amm 1amp 1anf 1aoe 1aoh 1aop 1aqb 1aqu
1aqz 1arb 1aru 1atg 1atl 1atz 1auo 1avw 1axn 1ay7 1ayl 1ayx 1azo 1b0b 1b0u 1b0y
1b16 1b2p 1b3a 1b4k 1b4v 1b5e 1b67 1b6a 1b6g 1b8o 1b8z 1b9o 1b9w 1bab 1bb1 1bbh
1bbz 1bd0 1bdm 1bdo 1bec 1beh 1ben 1bf4 1bf6 1bfd 1bfg 1bg2 1bg6 1bgc 1bgf 1bhp
1bi5 1bio 1bj7 1bk0 1bk7 1bkb 1bkf 1bkj 1bkp 1bkr 1bm8 1bn7 1bpi 1bqc 1bqk 1brt
1bs0 1bs9 1bsm 1bte 1btk 1bty 1bu7 1bu8 1bue 1bup 1bw9 1bx4 1bx7 1bxa 1bxo 1byi
1byq 1c02 1c0p 1c1d 1c1k 1c1l 1c24 1c3d 1c3p 1c3w 1c4q 1c52 1c5c 1c5e 1c75 1c7k
1c8c 1c90 1c9o 1cb0 1cc8 1ccw 1ccz 1cem 1ceq 1cex 1cf9 1cg5 1cgo 1chd 1cip 1cjc
1cjw 1cka 1cke 1cl8 1cmb 1cnv 1cnz 1co6 1cot 1cpq 1cqm 1cru 1cs1 1cse 1csh 1ctf
1ctj 1ctq 1cuo 1cv2 1cv8 1cvl 1cxc 1cxq 1cxy 1cy5 1cyd 1cyo 1czf 1czp 1d02 1d0c
1d0d 1d1q 1d2n 1d2s 1d3g 1d3v 1d4a 1d4o 1d4t 1d5t 1d7p 1d8w 1dbf 1dbg 1dbo 1dbw
1dc1 1dci 1dcs 1dd9 1ddw 1deu 1df4 1df7 1dfm 1dfu 1dg6 1dgf 1dgw 1dhn 1di6 1dif
1din 1dix 1dj0 1dk8 1dki 1dl2 1dlf 1dlw 1dmg 1dmh 1dnl 1dos 1doz 1dp7 1dps 1dpt
1dqg 1dqi 1dqs 1dqz 1ds1 1dsz 1dtd 1duv 1dvj 1dw9 1dxg 1dy5 1dyp 1dyq 1dys 1dz3
1dzk 1dzo 1e0w 1e19 1e29 1e2k 1e2w 1e30 1e43 1e4c 1e4m 1e58 1e5k 1e5m 1e5p 1e6b
```

1e6u 1e6w 1e7l 1e85 1e87 1e9g 1eaj 1ear 1eax 1eb6 1eca 1eco 1ecs 1edg 1edm 1edq
1eex 1eg9 1egu 1egw 1ej0 1ej8 1ejd 1ejg 1ek0 1ek6 1ekq 1elk 1elu 1elv 1elw 1emv
1en2 1enf 1eon 1ep0 1eq9 1eqj 1eqo 1erm 1erv 1erx 1erz 1es5 1es9 1et1 1etn 1eu1
1eu3 1euv 1euw 1evh 1evl 1ew0 1ew4 1ewf 1exm 1exr 1eye 1eyh 1eyv 1ezg 1ezm 1ezw
1f0i 1f0l 1f1e 1f2a 1f2t 1f3u 1f46 1f4p 1f5n 1f5v 1f60 1f6b 1f74 1f7d 1f7l 1f86
1f8e 1f94 1f9v 1f9z 1fas 1faz 1fbn 1fcq 1fcy 1fd3 1fdr 1fds 1fec 1fg7 1fgy 1fh0
1fhu 1fi2 1fiu 1fj2 1fjh 1fjj 1fk5 1fkj 1fl0 1flm 1flp 1flt 1fm0 1fmb 1fmk 1fn8
1fna 1fnc 1fnd 1fo8 1fp2 1fqt 1fr3 1fr7 1fs7 1fsg 1ft5 1ftk 1ftr 1fus 1fvg 1fvk
1fw9 1fx2 1fxd 1fxo 1fye 1fzq 1g0o 1g12 1g1t 1g24 1g2b 1g2n 1g2q 1g2r 1g2y 1g3p
1g4i 1g4y 1g57 1g5a 1g61 1g66 1g6g 1g6s 1g6u 1g6x 1g7a 1g8k 1g8q 1g9o 1ga6 1gai
1gbs 1gca 1gce 1gci 1gco 1gcq 1gcu 1gd0 1gd1 1gdj 1gdo 1gdv 1geg 1gg6 1gj7 1gk8
1gk9 1gkl 1gkm 1gmi 1gmu 1gmx 1gnl 1gof 1goi 1gp0 1gpe 1gpi 1gqv 1gso 1gte 1gtv
1guq 1gut 1gvp 1gx5 1h2r 1h4x 1h5q 1h61 1h6f 1h6h 1h6t 1h75 1h7n 1h7z 1h80 1h8d
1h96 1h97 1h98 1h99 1h9m 1hbn 1hbz 1hcl 1hcr 1hd2 1hdh 1hdo 1heu 1hfc 1hfe 1hfo
1hfu 1hg7 1hj9 1hka 1hlr 1hmt 1hnj 1hoz 1hp1 1hpm 1hq1 1hqk 1hqs 1htr 1hty 1hvb
1hw1 1hx0 1hx6 1hxi 1hxn 1hxr 1hyo 1hyv 1hz4 1hz6 1hzt 1i0d 1i0h 1i0r 1i0v 1i12
1i19 1i1j 1i1n 1i27 1i2s 1i2t 1i40 1i4f 1i4u 1i52 1i58 1i5g 1i60 1i6w 1i71 1i88
1i8o 1i9s 1ia8 1iab 1iat 1iby 1ic6 1icr 1id0 1ida 1ido 1ie9 1ifc 1ig5 1igq 1iho
1ihr 1ii5 1iib 1ij2 1ijq 1ijv 1ijy 1ikp 1im5 1in4 1inl 1io7 1iqq 1iqz 1ird 1irq
1isu 1it2 1itx 1iu8 1iua 1ixh 1j6z 1j77 1j79 1j83 1j8f 1j98 1j9b 1j9q 1ja9 1jat
1jay 1jb3 1jb9 1jbe 1jcl 1jd0 1jdr 1je0 1jek 1jer 1jet 1jf8 1jfb 1jfu 1jfx 1jg1
1jhd 1jhg 1jhj 1jix 1jjy 1jk3 1jke 1jks 1jkx 1jl0 1jl1 1jlj 1jlt 1jm0 1jnd 1jnr
1jo8 1jp4 1jpz 1jq5 1jqc 1jr8 1jrr 1jsr 1jvp 1jw9 1jx4 1jx6 1jy1 1jy2 1jye 1jyk
1jyr 1jz8 1jzg 1k0m 1k1e 1k20 1k3i 1k4g 1k4i 1k4v 1k55 1k6a 1k6f 1k6x 1k77 1k7c
1k8u 1k92 1k94 1ka1 1kaf 1kao 1kap 1kb0 1kcq 1kgd 1kgs 1kid 1kko 1koe 1koi 1kp6
1kpf 1kpt 1kq3 1kqf 1kqr 1ks9 1kth 1ktp 1kuh 1kv5 1kv7 1kv8 1kve 1kvl 1kwf 1kyp
1kzk 1l3k 1l6s 1l6x 1l7m 1lam 1lbu 1lcl 1lkk 1llp 1lmb 1lst 1luc 1m6p 1mba 1mct
1mdc 1mfa 1mfi 1mfm 1mgt 1mh1 1mjh 1mla 1mml 1mof 1mol 1moq 1mpg 1mrj 1mro 1mrp
1msi 1msk 1mty 1mug 1mun 1nar 1nbc 1ndd 1nfn 1nfp 1nif 1nkd 1nkr 1nls 1noa 1not
1nox 1npk 1nsc 1nul 1nwp 1nzy 1oaa 1onc 1opd 1orc 1osa 1pa2 1pcf 1pda 1pdo 1pef
1pen 1pgs 1phn 1php 1pin 1plc 1pmi 1poa 1ppn 1ppt 1psr 1ptf 1pym 1qau 1qb7 1qbz
1qcx 1qcz 1qd1 1qd9 1qdd 1qe3 1qf9 1qfm 1qft 1qg8 1qgi 1qgq 1qgv 1qgw 1qh4 1qh5
1qh8 1qhf 1qho 1qhq 1qhv 1qip 1qj4 1qjc 1qjd 1qjp 1qk5 1qkk 1qks 1ql0 1ql3 1qlw
1qmq 1qmv 1qnf 1qnj 1qnr 1qop 1qq4 1qq5 1qq9 1qqf 1qqq 1qre 1qrr 1qs1 1qsa 1qsg
1qst 1qtn 1qto 1qts 1qtw 1qu9 1qup 1qus 1ra9 1rb9 1rcf 1rge 1rhs 1rie 1rro 1rzl
1sbp 1sgp 1sgt 1slu 1smd 1sml 1stn 1svf 1svy 1swu 1t1d 1tax 1tc1 1tca 1ten 1tfe
1tgs 1tgx 1thf 1thm 1thv 1thx 1tif 1tml 1toa 1tph 1ttb 1tud 1tx4 1tyv 1uae 1ubp
1uch 1ugi 1uro 1ush 1ute 1utg 1uxy 1vca 1vcc 1vfr 1vfy 1vhh 1vie 1vjs 1vns 1vsr
1wab 1wap 1wer 1wfb 1whi 1xik 1xjo 1xnb 1xwl 1xyz 1yac 1ycc 1yge 1yna 1ypc 1ytb
1yve 1zin 256b 2a0b 2act 2acy 2ahj 2arc 2ayh 2baa 2bbk 2bc2 2bce 2bop 2btc 2bvw
2cba 2cbp 2ccy 2cpg 2cpl 2cpp 2ctc 2cth 2cua 2cy3 2cyp 2dpm 2dri 2end 2eng 2erl
2fdn 2gar 2gdm 2hbg 2hft 2hlc 2hmz 2igd 2ilk 2knt 2lis 2ltn 2mcm 2mhr 2msb 2myr
2nac 2nlr 2por 2pth 2pvb 2qwc 2rhe 2rn2 2sak 2sn3 2sns 2spc 2tgi 2tnf 2tps 2trx
3bto 3c2c 3cao 3chb 3chy 3cla 3cyr 3dfr 3ebx 3eip 3ezm 3grs 3hts 3lzt 3nul 3pro
3pte 3pvi 3pyp 3sdh 3seb 3sil 3std 3vub 451c 4eug 4lzt 4pga 4uag 4ubp 4xis 5cyt
5hpg 5icb 5nul 5p21 5pal 5rub 6cel 6gsv 6rlx 7a3h 7atj 7fd1 7odc 7rsa 8abp 8dfr
8ruc 8tln 9wga

## Geometrical requirements for test cases

Each table represents one type of contact. Within each table, sets of alternative geometry definitions are provided. Satisfying any one set of geometry definitions within a table is sufficient to meet that contact requirement.

### Chorismate mutase

Five contacts are required: two planar guanidino-carboxylate contacts to arginine, one lysine and one glutamine contact to the ether oxygen of the breaking bond, and one glutamate contact to the hydroxyl group. Atom names follow the naming scheme of ref. 8.

Contact: Arginine to C11 carboxylate

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **ARG** | PSEUDO_ATOM | PSI1, equidistant between C11A & C11B | | | | | |
| | PSEUDO_ATOM | PSI2, equidistant between NH1 & NH2 | | | | | |
| | DISTANCE | PSI2 | PSI1 | | | 2.5 | 3.4 |
| | ANGLE | NH1 | PSI2 | PSI1 | | 69.0 | 111.0 |
| | TORSION | CZ | NH1 | PSI2 | PSI1 | 159.0 | 201.0 |
| | ANGLE | PSI2 | PSI1 | C11A | | 69.0 | 111.0 |
| | TORSION | NH1 | PSI2 | PSI1 | C11A | -21.0 | 21.0 |
| | TORSION | PSI2 | PSI1 | C11A | C11 | 159.0 | 201.0 |
| | DISTANCE | NH1 | C11A | | | 2.5 | 3.4 |
| | DISTANCE | NH2 | C11B | | | 2.5 | 3.4 |
| **ARG** | PSEUDO_ATOM | PSI1, equidistant between C11A & C11B | | | | | |
| | PSEUDO_ATOM | PSI2, equidistant between NH1 & NH2 | | | | | |
| | DISTANCE | PSI2 | PSI1 | | | 2.5 | 3.4 |
| | ANGLE | NH2 | PSI2 | PSI1 | | 69.0 | 111.0 |
| | TORSION | CZ | NH2 | PSI2 | PSI1 | 159.0 | 201.0 |
| | ANGLE | PSI2 | PSI1 | C11A | | 69.0 | 111.0 |
| | TORSION | NH2 | PSI2 | PSI1 | C11A | -21.0 | 21.0 |
| | TORSION | PSI2 | PSI1 | C11A | C11 | 159.0 | 201.0 |
| | DISTANCE | NH2 | C11A | | | 2.5 | 3.4 |
| | DISTANCE | NH1 | C11B | | | 2.5 | 3.4 |
| **ARG** | PSEUDO_ATOM | PSI1, equidistant between C11A & C11B | | | | | |
| | PSEUDO_ATOM | PSI2, equidistant between NH1 & NE | | | | | |
| | DISTANCE | PSI2 | PSI1 | | | 2.5 | 3.4 |
| | ANGLE | NH1 | PSI2 | PSI1 | | 69.0 | 111.0 |
| | TORSION | CZ | NH1 | PSI2 | PSI1 | 159.0 | 201.0 |
| | ANGLE | PSI2 | PSI1 | C11A | | 69.0 | 111.0 |
| | TORSION | NH1 | PSI2 | PSI1 | C11A | -21.0 | 21.0 |
| | TORSION | PSI2 | PSI1 | C11A | C11 | 159.0 | 201.0 |
| | DISTANCE | NE | C11B | | | 2.5 | 3.4 |
| | DISTANCE | NH1 | C11A | | | 2.5 | 3.4 |
| **ARG** | PSEUDO_ATOM | PSI1, equidistant between C11A & C11B | | | | | |
| | PSEUDO_ATOM | PSI2, equidistant between NH1 & NE | | | | | |
| | DISTANCE | PSI2 | PSI1 | | | 2.5 | 3.4 |
| | ANGLE | NH1 | PSI2 | PSI1 | | 69.0 | 111.0 |
| | TORSION | CZ | NH1 | PSI2 | PSI1 | 159.0 | 201.0 |
| | ANGLE | PSI2 | PSI1 | C11A | | 69.0 | 111.0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | TORSION | NH1 | PSI2 | PSI1 | C11A | 159.0 | 201.0 |
| | TORSION | PSI2 | PSI1 | C11A | C11 | 159.0 | 201.0 |
| | DISTANCE | NE | C11A | | | 2.5 | 3.4 |
| | DISTANCE | NH1 | C11B | | | 2.5 | 3.4 |

Contact: Arginine to C10 carboxylate

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **ARG** | PSEUDO_ATOM | PSI1, equidistant between C10A & C10B | | | | | |
| | PSEUDO_ATOM | PSi2, equidistant between NH1 & NH2 | | | | | |
| | DISTANCE | PSI2 | PSI1 | | | 2.5 | 3.4 |
| | ANGLE | NH1 | PSI2 | PSI1 | | 69.0 | 111.0 |
| | TORSION | CZ | NH1 | PSI2 | PSI1 | 159.0 | 201.0 |
| | ANGLE | PSI2 | PSI1 | C10A | | 69.0 | 111.0 |
| | TORSION | NH1 | PSI2 | PSI1 | C10A | -21.0 | 21.0 |
| | TORSION | PSI2 | PSI1 | C10A | C10 | 159.0 | 201.0 |
| | DISTANCE | NH1 | C10A | | | 2.5 | 3.4 |
| | DISTANCE | NH2 | C10B | | | 2.5 | 3.4 |
| **ARG** | PSEUDO_ATOM | PSI1, equidistant between C10A & C10B | | | | | |
| | PSEUDO_ATOM | PSI2, equidistant between NH1 & NH2 | | | | | |
| | DISTANCE | PSI2 | PSI1 | | | 2.5 | 3.4 |
| | ANGLE | NH1 | PSI2 | PSI1 | | 69.0 | 111.0 |
| | TORSION | CZ | NH1 | PSI2 | PSI1 | 159.0 | 201.0 |
| | ANGLE | PSI2 | PSI1 | C10A | | 69.0 | 111.0 |
| | TORSION | NH1 | PSI2 | PSI1 | C10A | 159.0 | 201.0 |
| | TORSION | PSI2 | PSI1 | C10A | C10 | 159.0 | 201.0 |
| | DISTANCE | NH1 | C10B | | | 2.5 | 3.4 |
| | DISTANCE | NH2 | C10A | | | 2.5 | 3.4 |
| **ARG** | PSEUDO_ATOM | PSI1, equidistant between C10A & C10B | | | | | |
| | PSEUDO_ATOM | PSI2, equidistant between NH1 & NE | | | | | |
| | DISTANCE | PSI2 | PSI1 | | | 2.5 | 3.4 |
| | ANGLE | NH1 | PSI2 | PSI1 | | 69.0 | 111.0 |
| | TORSION | CZ | NH1 | PSI2 | PSI1 | 159.0 | 201.0 |
| | ANGLE | PSI2 | PSI1 | C10A | | 69.0 | 111.0 |
| | TORSION | NH1 | PSI2 | PSI1 | C10A | -21.0 | 21.0 |
| | TORSION | PSI2 | PSI1 | C10A | C10 | 159.0 | 201.0 |
| | DISTANCE | NE | C10B | | | 2.5 | 3.4 |
| | DISTANCE | NH1 | C10A | | | 2.5 | 3.4 |
| **ARG** | PSEUDO_ATOM | PSI1, equidistant between C10A & C10B | | | | | |
| | PSEUDO_ATOM | PSI2, equidistant between NH1 & NE | | | | | |
| | DISTANCE | PSI2 | PSI1 | | | 2.5 | 3.4 |
| | ANGLE | NH1 | PSI2 | PSI1 | | 69.0 | 111.0 |
| | TORSION | CZ | NH1 | PSI2 | PSI1 | 159.0 | 201.0 |
| | ANGLE | PSI2 | PSI1 | C10A | | 69.0 | 111.0 |
| | TORSION | NH1 | PSI2 | PSI1 | C10A | 159.0 | 201.0 |
| | TORSION | PSI2 | PSI1 | C10A | C10 | 159.0 | 201.0 |
| | DISTANCE | NE | C10A | | | 2.5 | 3.4 |
| | DISTANCE | NH1 | C10B | | | 2.5 | 3.4 |

Contact: Lysine to O7

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **LYS** | DISTANCE | O7 | NZ | | | 2.5 | 3.4 |
| | ANGLE | C8 | O7 | NZ | | 89.5 | 159.5 |
| | TORSION | C9 | C8 | O7 | NZ | 150.0 | 210.0 |
| **LYS** | DISTANCE | O7 | NZ | | | 2.5 | 3.4 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ANGLE | C8 | O7 | NZ | | 89.5 | 159.5 |
| | TORSION | C9 | C8 | O7 | NZ | 30.0 | 90.0 |

Contact: Glutamine to O7

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **GLN** | DISTANCE | O7 | NE2 | | | 2.5 | 3.4 |
| | ANGLE | C8 | O7 | NE2 | | 89.5 | 129.5 |
| | TORSION | C9 | C8 | O7 | NE2 | 150.0 | 210.0 |
| | ANGLE | O7 | 1HE2 | NE2 | | 150.0 | 180.0 |
| **GLN** | DISTANCE | O7 | NE2 | | | 2.5 | 3.4 |
| | ANGLE | C8 | O7 | NE2 | | 89.5 | 129.5 |
| | TORSION | C9 | C8 | O7 | NE2 | 30.0 | 90.0 |
| | ANGLE | O7 | 1HE2 | NE2 | | 150.0 | 180.0 |
| **GLN** | DISTANCE | O7 | NE2 | | | 2.5 | 3.4 |
| | ANGLE | C8 | O7 | NE2 | | 89.5 | 129.5 |
| | TORSION | C9 | C8 | O7 | NE2 | 150.0 | 210.0 |
| | ANGLE | O7 | 2HE2 | NE2 | | 150.0 | 180.0 |
| **GLN** | DISTANCE | O7 | NE2 | | | 2.5 | 3.4 |
| | ANGLE | C8 | O7 | NE2 | | 89.5 | 129.5 |
| | TORSION | C9 | C8 | O7 | NE2 | 30.0 | 90.0 |
| | ANGLE | O7 | 2HE2 | NE2 | | 150.0 | 180.0 |

Contact: Glutamate to O4 hydroxyl group

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **GLU** | DISTANCE | O4 | OE1 | | | 2.5 | 3.4 |
| | ANGLE | O4 | O4H | OE1 | | 150.0 | 180.0 |
| **GLU** | DISTANCE | O4 | OE2 | | | 2.5 | 3.4 |
| | ANGLE | O4 | O4H | OE2 | | 150.0 | 180.0 |

**Streptavidin-biotin**

Five contacts are required: Serine, asparagine, and tyrosine hydrogen bond donors to O3 and aspartate and serine hydrogen bond acceptors at N1 and N2, respectively. Atom names are from pdb code 1mk5.

Contact: Serine to O3

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **SER** | DISTANCE | O3 | OG | | | 2.6 | 3.3 |
| | ANGLE | O3 | HG | OG | | 150.0 | 180.0 |

Contact: Asparagine to O3

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **ASN** | DISTANCE | O3 | ND2 | | | 2.6 | 3.3 |
| | ANGLE | O3 | 1HD2 | ND2 | | 150.0 | 180.0 |
| **ASN** | DISTANCE | O3 | ND2 | | | 2.6 | 3.3 |
| | ANGLE | O3 | 2HD2 | ND2 | | 150.0 | 180.0 |

Contact: Tyrosine to O3

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **TYR** | DISTANCE | O3 | OH | | | 2.6 | 3.3 |
| | ANGLE | O3 | HH | OH | | 150.0 | 180.0 |

Contact: Aspartate to N1

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **ASP** | DISTANCE | N1 | OD1 | | | 2.6 | 3.3 |
| | ANGLE | N1 | H1 | OD1 | | 150.0 | 180.0 |
| **ASP** | DISTANCE | N1 | OD2 | | | 2.6 | 3.3 |
| | ANGLE | N1 | H1 | OD2 | | 150.0 | 180.0 |

Contact: Serine to N2

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **SER** | DISTANCE | N2 | OG | | | 2.6 | 3.3 |
| | ANGLE | N2 | 2H2 | OG | | 150.0 | 180.0 |

**Triosephosphate isomerase**

Three contacts are required: Lysine, histidine, and glutamate contacts as shown in Figure 1. Atom names are from PDB code 1ney.

Contact: Lysine to O1/O2

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **LYS** | DISTANCE | O2 | NZ | | | 2.8 | 3.2 |
| | DISTANCE | O1 | NZ | | | 2.8 | 3.2 |
| | ANGLE | O2 | 1HZ | NZ | | 140.0 | 180.0 |
| | TORSION | C1 | C2 | O2 | NZ | 60.0 | 120.0 |
| **LYS** | DISTANCE | O2 | NZ | | | 2.8 | 3.2 |
| | DISTANCE | O1 | NZ | | | 2.8 | 3.2 |
| | ANGLE | O2 | 2HZ | NZ | | 140.0 | 180.0 |
| | TORSION | C1 | C2 | O2 | NZ | 60.0 | 120.0 |
| **LYS** | DISTANCE | O2 | NZ | | | 2.8 | 3.2 |
| | DISTANCE | O1 | NZ | | | 2.8 | 3.2 |
| | ANGLE | O2 | 3HZ | NZ | | 140.0 | 180.0 |
| | TORSION | C1 | C2 | O2 | NZ | 60.0 | 120.0 |

Contact: Histidine to O2/O3

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **HIS** | PSEUDO_ATOM | PSA, equidistant between O2 & O3 | | | | | |
| | DISTANCE | PSA | NE2 | | | 2.3 | 2.7 |
| | ANGLE | PSA | HE2 | NE2 | | 140.0 | 180.0 |
| | TORSION | PSA | NE2 | CE1 | ND1 | 160.0 | 200.0 |
| | ANGLE | O2 | PSA | NE2 | | 70.0 | 110.0 |
| | TORSION | O2 | PSA | NE2 | CE1 | 0.0 | 360.0 |
| | TORSION | C2 | O2 | PSA | NE2 | 160.0 | 200.0 |
| **HIS** | PSEUDO_ATOM | PSA, equidistant between O2 & O3 | | | | | |
| | DISTANCE | PSA | ND1 | | | 2.3 | 2.7 |
| | ANGLE | PSA | HD1 | ND1 | | 140.0 | 180.0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | TORSION | PSA | ND1 | CE1 | NE2 | 160.0 | 200.0 |
| | ANGLE | O2 | PSA | ND1 | | 70.0 | 110.0 |
| | TORSION | O2 | PSA | ND1 | CE1 | 0.0 | 360.0 |
| | TORSION | C2 | O2 | PSA | ND1 | 160.0 | 200.0 |

Contact: Glutamate to O2/O3

| residue | type | atom1 | atom2 | atom3 | atom4 | min | max |
|---|---|---|---|---|---|---|---|
| **GLU** | PSEUDO_ATOM | PSA, equidistant between O2 & O3 | | | | | |
| | DISTANCE | PSA | CD | | | 3.2 | 3.6 |
| | ANGLE | C3 | C2 | CD | | 60.0 | 120.0 |
| | TORSION | C2 | C3 | OE1 | CD | 60.0 | 120.0 |
| | | | | | | | |
| **GLU** | PSEUDO_ATOM | PSA, equidistant between O2 & O3 | | | | | |
| | DISTANCE | PSA | CD | | | 3.2 | 3.6 |
| | ANGLE | C3 | C2 | CD | | 60.0 | 120.0 |
| | TORSION | C2 | C3 | OE2 | CD | 60.0 | 120.0 |

## Variation of contact geometry for targeted placement

Each type of placement was performed for every rotamer of the indicated amino acid in every designed position. For each type of placement, every distance, angle, and torsion was varied by the range and step size indicated, in combination with all of the other distance, angle, and torsional variations for that placement type.

### Chorismate mutase

Placement: carboxylate (C11A, C11B) from Arginine (NH1, NH2)

| type | atom1 | atom2 | atom3 | atom4 | min | max | step |
|---|---|---|---|---|---|---|---|
| PSEUDO_ATOM | PSI1, equidistant between C11A & C11B | | | | | | |
| PSEUDO_ATOM | PSI2, equidistant between NH1 & NH2 | | | | | | |
| DISTANCE | PSI2 | PSI1 | | | 2.8 | 3.4 | 0.2 |
| ANGLE | NH1 | PSI2 | PSI1 | | 80 | 100 | 10 |
| TORSION | CZ | NH1 | PSI2 | PSI1 | 170 | 190 | 10 |
| ANGLE | PSI2 | PSI1 | C11A | | 80 | 100 | 10 |
| TORSION | NH1 | PSI2 | PSI1 | C11A | -10 | 10 | 10 |
| TORSION | PSI2 | PSI1 | C11A | C11 | 170 | 190 | 10 |

Placement: carboxylate (C11A, C11B) from Arginine (NH2, NH1)

| type | atom1 | atom2 | atom3 | atom4 | min | max | step |
|---|---|---|---|---|---|---|---|
| PSEUDO_ATOM | PSI1, equidistant between C11A & C11B | | | | | | |
| PSEUDO_ATOM | PSI2, equidistant between NH1 & NH2 | | | | | | |
| DISTANCE | PSI2 | PSI1 | | | 2.8 | 3.4 | 0.2 |
| ANGLE | NH1 | PSI2 | PSI1 | | 80 | 100 | 10 |
| TORSION | CZ | NH1 | PSI2 | PSI1 | 170 | 190 | 10 |
| ANGLE | PSI2 | PSI1 | C11A | | 80 | 100 | 10 |
| TORSION | NH1 | PSI2 | PSI1 | C11A | 170 | 190 | 10 |
| TORSION | PSI2 | PSI1 | C11A | C11 | 170 | 190 | 10 |

Placement: carboxylate (C11A, C11B) from Arginine (NE, NH1)

| type | atom1 | atom2 | atom3 | atom4 | min | max | step |
|---|---|---|---|---|---|---|---|
| PSEUDO_ATOM | PSI1, equidistant between C11A & C11B | | | | | | |
| PSEUDO_ATOM | PSI2, equidistant between NE & NH1 | | | | | | |
| DISTANCE | PSI2 | PSI1 | | | 2.8 | 3.4 | 0.2 |
| ANGLE | NH1 | PSI2 | PSI1 | | 80 | 100 | 10 |
| TORSION | CZ | NH1 | PSI2 | PSI1 | 170 | 190 | 10 |
| ANGLE | PSI2 | PSI1 | C11A | | 80 | 100 | 10 |
| TORSION | NH1 | PSI2 | PSI1 | C11A | 170 | 190 | 10 |
| TORSION | PSI2 | PSI1 | C11A | C11 | 170 | 190 | 10 |

Placement: carboxylate (C11A, C11B) from Arginine (NH1, NE)

| type | atom1 | atom2 | atom3 | atom4 | min | max | step |
|---|---|---|---|---|---|---|---|
| PSEUDO_ATOM | PSI1, equidistant between C11A & C11B | | | | | | |
| PSEUDO_ATOM | PSI2, equidistant between NE & NH1 | | | | | | |
| DISTANCE | PSI2 | PSI1 | | | 2.8 | 3.4 | 0.2 |
| ANGLE | NH1 | PSI2 | PSI1 | | 80 | 100 | 10 |
| TORSION | CZ | NH1 | PSI2 | PSI1 | 170 | 190 | 10 |
| ANGLE | PSI2 | PSI1 | C11A | | 80 | 100 | 10 |
| TORSION | NH1 | PSI2 | PSI1 | C11A | -10 | 10 | 10 |
| TORSION | PSI2 | PSI1 | C11A | C11 | 170 | 190 | 10 |

## Streptavidin-Biotin

Placement: N1 from Aspartate (OD2)

| type | atom1 | atom2 | atom3 | atom4 | min | max | step |
|---|---|---|---|---|---|---|---|
| DISTANCE | OD2 | N1 | | | 2.8 | 2.8 | 0 |
| ANGLE | CG | OD2 | H1 | | 120 | 120 | 0 |
| TORSION | CG | OD1 | OD2 | H1 | 160 | 200 | 10 |
| ANGLE | OD2 | H1 | N1 | | 160 | 180 | 10 |
| TORSION | OD1 | OD2 | H1 | C3 | 210 | 270 | 10 |

Placement: N1 from Aspartate (OD1)

| type | atom1 | atom2 | atom3 | atom4 | min | max | step |
|---|---|---|---|---|---|---|---|
| DISTANCE | OD1 | N1 | | | 2.8 | 2.8 | 0 |
| ANGLE | CG | OD1 | N1 | | 120 | 120 | 0 |
| TORSION | CG | OD2 | OD1 | H1 | 160 | 200 | 10 |
| ANGLE | OD1 | H1 | N1 | | 160 | 180 | 10 |
| TORSION | OD2 | OD1 | H1 | C3 | 210 | 270 | 10 |

## Triosephosphate isomerase

Placement: O2/O3 from Histidine (ND1)

| type | atom1 | atom2 | atom3 | atom4 | min | max | step |
|---|---|---|---|---|---|---|---|
| PSEUDO_ATOM | PSA, equidistant between O2 & O3 | | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| DISTANCE | ND1 | PSA | | | 2.3 | 2.7 | 0.1 |
| ANGLE | ND1 | HD1 | PSA | | 140 | 180 | 20 |
| TORSION | NE2 | CE1 | ND1 | PSA | 160 | 200 | 20 |
| ANGLE | ND1 | PSA | O2 | | 70 | 110 | 20 |
| TORSION | CE1 | ND1 | PSA | O2 | 0 | 360 | 20 |
| TORSION | ND1 | PSA | O2 | C2 | 160 | 200 | 20 |

Placement: O2/O3 from Histidine (NE2)

| type | atom1 | atom2 | atom3 | atom4 | min | max | step |
|---|---|---|---|---|---|---|---|
| PSEUDO_ATOM | PSA, equidistant between O2 & O3 | | | | | | |
| DISTANCE | NE2 | PSA | | | 2.3 | 2.7 | 0.1 |
| ANGLE | NE2 | HE2 | PSA | | 140 | 180 | 20 |
| TORSION | ND1 | CE1 | NE2 | PSA | 160 | 200 | 20 |
| ANGLE | NE2 | PSA | O2 | | 70 | 110 | 20 |
| TORSION | CE1 | NE2 | PSA | O2 | 0 | 360 | 20 |
| TORSION | NE2 | PSA | O2 | C2 | 160 | 200 | 20 |

# Atomic charges Used

## HF/6-31G* transition state structure, chorismate mutase (8)

| | |
|---|---|
| O7 | -0.59 |
| C8 | 0.31 |
| C9 | -0.56 |
| C9H1 | 0.19 |
| C9H2 | 0.16 |
| C10 | 0.86 |
| C10A | -0.84 |
| C10B | -0.84 |
| C5 | 0.25 |
| C6 | -0.62 |
| C6H | 0.20 |
| C1 | 0.43 |
| C2 | -0.42 |
| C3 | -0.03 |
| C4 | 0.40 |
| C4H | 0.04 |
| O4 | -0.79 |
| O4H | 0.40 |
| C3H | 0.07 |
| C2H | 0.18 |
| C11 | 0.80 |
| C11B | -0.81 |
| C11A | -0.83 |
| C5H | 0.04 |

## Biotin

| | |
|---|---|
| C11 | 0.88 |
| O11 | -0.84 |
| O12 | -0.84 |
| C10 | -0.22 |
| 1H10 | 0.02 |
| 2H10 | 0.04 |
| C9 | -0.02 |
| 1H9 | -0.03 |
| 2H9 | 0.03 |
| C8 | 0.04 |
| 1H8 | 0.02 |
| 2H8 | 0.02 |
| C7 | 0.08 |
| 1H7 | -0.01 |
| 2H7 | 0.00 |
| C2 | -0.03 |
| 1H2 | 0.05 |
| S1 | -0.33 |
| C6 | 0.09 |
| 1H6 | 0.00 |
| 2H6 | 0.06 |
| C5 | 0.16 |
| H5 | 0.01 |
| N1 | -0.64 |
| H1 | 0.35 |
| C3 | 0.78 |
| O3 | -0.67 |
| N2 | -0.63 |
| 2H2 | 0.34 |
| C4 | 0.31 |
| H4 | -0.02 |

**Dihydroxyacetone phosphate**

| | |
|---|---|
| O1 | -0.42 |
| N2 | -0.22 |
| C3 | 0.14 |
| H3 | 0.19 |
| C4 | -0.05 |
| C5 | -0.11 |
| H5 | 0.19 |
| C6 | 0.24 |
| C7 | -0.13 |
| H7 | 0.18 |
| C8 | -0.11 |
| H8 | 0.17 |
| C9 | 0.37 |
| N6 | 0.42 |
| O61 | -0.44 |
| O62 | -0.43 |

**References**


1. Ponder JW, Richards FM (1987) *J Mol Biol* 193:775-791.

2. Dunbrack RL Jr., Cohen FE (1997) *Prot Sci* 6:1661-1681.

3. Lovell SC, Word JM, Richardson JS, Richardson DC (2000) *Proteins* 40:389-408.

4. Shetty PR, de Bakker PIW, DePristo MA, Blundell TL (2003) *Prot Eng* 16:963-969.

5. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) *J Mol Biol* 285:1735-1747.

6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) *J Mol Biol* 215: 403–410.

7. Mayo SL, Olafson BD, Goddard WA (1990) *J Phys Chem* 94:8897-8909.

8. Wiest O, Houk KN (1994) *J Org Chem* 59:7582-7584.